ERDC/CERL TR-01-46

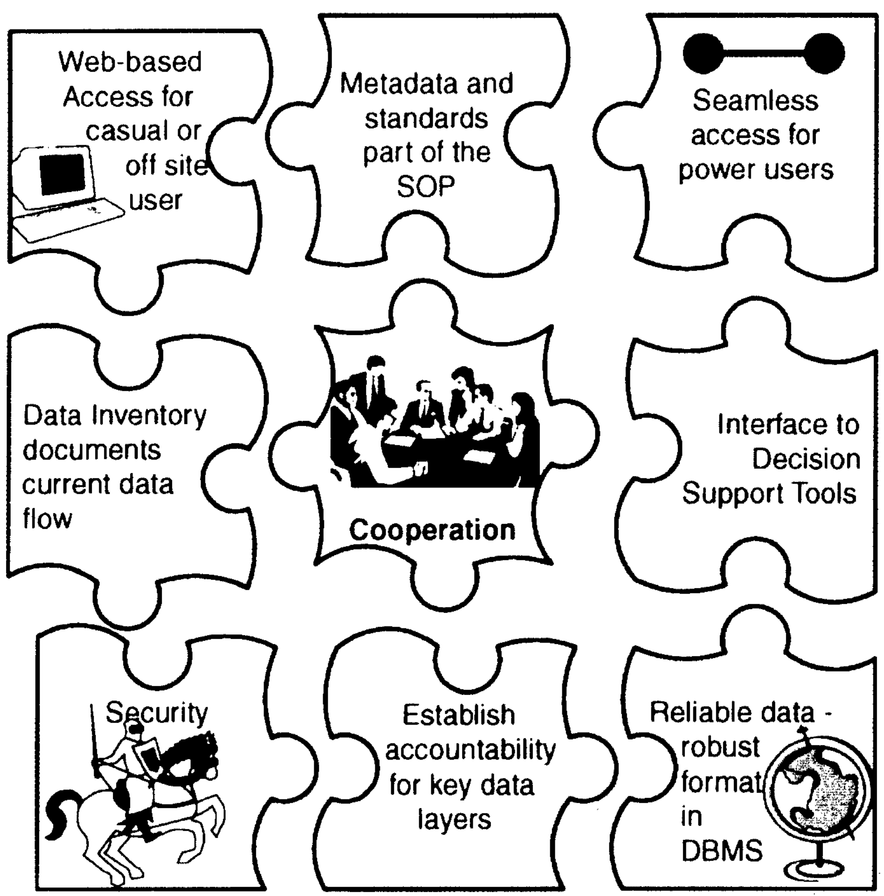
Construction Engineering Research Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

Geospatial Data Enterprise Repository: A Report on the Prototype for Fort Hood, Texas

Marilyn Ruiz, Dawn Morrison, David Bouwman, Kevin McNinch, Frank Schreiner, and Shari Forbes

May 2001



Foreword

This study was conducted for Headquarters, U.S. Army Corps of Engineers under project number 212040622720A917, Congressional; Work Unit BH0, "Data Repository and Web Mapping." The technical monitor was Dr. Alan Moore, CEERD-CV-T.

The work was performed by the Land and Heritage Conservation Branch (CN-C) of the Installations Division (CN) Construction Engineering Research Laboratory (CERL). The CERL Principal Investigator was Dr. Marilyn O. Ruiz. Part of this work was done by Dawn Morrison, University of Illinois; David Bouwman, Kevin McNinch, and Frank Schreiner of Pacific Meridian Resources; and Shari Forbes, Southwest Texas State University. The technical editor was Gloria J. Wienke, Information Technology Laboratory. Stephen E. Hodapp is Chief, CEERD-CN-C, and Dr. John T. Bandy is Chief, CEERD-CN. The associated Technical Director was Dr. Alan W. Moore, CEERD-CV-T. The Acting Director of CERL is Dr. Alan W. Moore.

CERL is an element of the U.S. Army Engineer Research and Development Center (ERDC), U.S. Army Corps of Engineers. The Director of ERDC is Dr. James R. Houston and the Deputy to the Commander is A.J. Roberto, Jr.

Executive Summary

Geographic Information System (GIS) data management options have expanded rapidly over the past several years. Software and hardware advances have provided better network access to spatial data, allowed more complex geospatial data models, and have made the integration of GIS and database management systems a reality.

At Fort Hood, Texas, GIS technology has been used extensively for military land management. Though their current technology is mature in many ways, there remain several issues related to geospatial data management that hinder efficiency. Primary issues included duplication of data themes in various versions across offices, uncertainty about accountability for the content of key data themes, turn over of GIS staff with a subsequent loss of institutional knowledge, and data requests taking time away from other duties.

This report outlines a proposed plan for a more centralized geospatial data repository, the Data Enterprise Repository (DER), to meet the geospatial data requirements of the installation's community. The core users of the system are the Environmental Division office, Cultural Resource Management Team, the Natural Resource Branch, and the Integrated Training Area Management office. The DER will also benefit those who need to view maps from the system but do not use a GIS, and will enable people offsite, who are performing work at the installation, to get better access to the data they require for their work.

The DER prototype project involved several activities. These included a data inventory of the geospatial data resident in the core GIS group workstations, discussions with core GIS users to better understand the issues, an assessment of relevant emerging technologies, development of a web interface to better structure geospatial data management practices, and a security assessment to determine the implications of potential security threats. This report describes the results of those activities.

This report does not cover an evaluation of the implementation process, because that has not yet occurred. The framework and options described in this report need to be discussed and adjusted by the core GIS users at the installation before the implementation can begin. Many of the issues raised in this study are com-

mon to many places, not just military installations, but for all of the sites where GIS operations have become fairly dispersed across an organization. Each organization brings its own institutional climate to the table of such discussions. The site-specific needs of the institution can be blended harmoniously with the technology that can allow it to operate more efficiently, but the mere presence of the technology will not guarantee that this efficiency is achieved.

Contents

List of Figures and Tables

Figures

Tables

1 Introduction

Background

The past 15 years have seen the rapid development and proliferation of networked information systems. The U.S. Army Corps of Engineers (USACE) has participated in this development through projects that promote network-based data sharing, communications, and systems access for users at dispersed locations. The USACE's Land Management System (LMS) initiative calls for extensive use of networked systems to connect LMS users to each other and to allow users access to the LMS components (Goran et al. 1999). An essential part of the total system is a delivery mechanism for data. The Data Enterprise Repository (DER) is designed to address LMS data requirements in the specific context of a military demonstration site. The current project focuses on an evaluation of technical options for a shared data repository with a strong geospatial component and the implementation issues that result from putting such a system in place at Fort Hood, Texas, an LMS demonstration site.

At root, the DER is a network-accessible repository of natural resource data. The prototype repository designed for LMS facilitates access to diverse land management datasets located at the installation. Much of the critical information used for land management decisions is stored as digital geospatial data sets, such as digital maps, satellite and aerial images, elevation models, and extensive relational databases. The data come from a variety of sources, and are generally in a state of flux, as new data sets are collected and existing data are updated. The data are used for a diverse range of studies and land management activities, such as those concerned with protection of threatened and endangered species, coordination of digging and construction permits, long term ecological monitoring, and assessment of training impacts. A multitude of participants, users, software, hardware, locations, and needs is the norm.

A number of issues illustrate the need for the DER at the installation:

1. The multiple complexities of data volume, Geographic Information Systems (GIS) data formats, multiple shared servers, and physical distance among data users have been impediments to seamless sharing of geospatial information and re-

lated databases. Currently, numerous versions of data with similar titles exist on various computers.

2. The current method for data sharing requires tapping into the knowledge of a set of key individuals. Responding to requests for data can be an impediment to fulfilling other critical functions. The method does not function when personnel are absent. A more automated, coherent method of data access would alleviate the pressure on individuals and provide a more efficient and effective way to share data.
3. Accountability for key data layers is not clearly established and documented.
4. There is no coherent "front end" so that the user sees a common, well-documented database. Each user's view depends on length of service, technical expertise, and access to technology.

The repository design seeks to provide a common focus for data collection, archiving, and access, thus eliminating the need for each location to create disparate access methods to this complex tapestry of data options. This will help to ensure the long term and widespread usefulness of the information used for land management decisions, and will protect the often extensive investment in data development common to natural resource management activities.

Objectives

The following technical objectives guided the prototype development and guided the decisions and path of system development.

- Provide an organizational framework for storing existing data
- Create a user interface to search for and locate data (catalog)
- Provide help to users in the evaluation of the usefulness of the data (metadata)
- Provide data management guidelines to data creators
- Assess need for and propose a plan for data security
- Allow access to data in the form needed at the point where it is required.

The Physical and Institutional Context for the DER

Fort Hood is an excellent location to assess the institutional and technological issues related to geospatial data management and distribution. Fort Hood, located in central Texas, is considered the Army's premier installation for the training and deployment of heavy forces. Established in 1942, and covering 340

square miles (217,337 acres), the installation is the only United States military base capable of accommodating two Armored Divisions: the 1st Cavalry Division and the 4th Infantry Division. Additional military units residing at Fort Hood include: Headquarters Command III Corps, 3rd Personnel Group, 3rd Signal Brigade, 3rd Air Support Operations Group, 13th Corps Support Command, 13th Finance Group, 21st CAV Brigade, 89th Military Police Brigade, 504th Military Intelligence Brigade, the Dental Activity (DENTAC), the Medical Support Activity (MEDDAC), and the Test and Experimentation Command (TEXCOM). The installation also serves as a training facility for the 49th Armored Division of the National Guard, as well as various other units.

Fort Hood is located on the northern edge of Texas Hill Country and on the eastern edge of the Lampasas Cut Plains region. This provides the installation with topography of rolling hills, shallow soils, small stream valleys, ridge-forming mesas, and a combination of woodlands and prairies. The installation also contains several lakes, nonjurisdictional wetlands, and a network of Karst features. There are six soil associations found within the base, ranging from limestone and shale to sandy clay, which support a variety of vegetation in the form of trees, shrubs, and grasses, along with one of the rarest shrubs in North America — the Texabama Croton. The climate of the base is semi-arid continental with mild winters and hot, dry summers.

The base is by far the most influential institution in the region. The total population served by the installation is approximately 218,003, including active duty soldiers, family members, civilian employees, and retirees, survivors, and their family members. Land use is comprised of approximately 10 percent improved land found in the cantonment area, and 90 percent unimproved land. Roughly 65 percent of the land is used for military training operations including maneuver, live fire, and impact areas. In addition, the land is used by local ranchers for a limited number of free-range cattle and by local residents for a variety of recreational activities.

A wide variety of birds, fish, reptiles, amphibians, and mammals make their home on the installation. Currently, three species of birds — the Golden-cheeked Warbler, the Black-capped Vireo, and the Bald Eagle — are listed as endangered species, and several programs and initiatives for conservation and preservation are in place to protect these species. Conversely, many species are managed as a recreational resource, such as the white-tailed deer and a variety of fish, while Brown-headed Cowbirds, considered parasites to the Black-capped Vireos, are actively trapped and removed from the base.

The installation adheres to a solid environmental ethic that seeks to minimize the impact of military training on the environment. The base's commitment to the environment is expressed in their vision statement for the environment:

> "Fort Hood recognizes the importance of instituting a sound environmental management plan, incorporating the four pillars [i.e., compliance, restoration, prevention and conservation] of environmental stewardship into daily operations, consistent with the accomplishment of the military mission. Fort Hood intends to be the leader in the environmental arena as the Army moves into the 21st Century" (Directorate of Public Works 2000).

The primary responsibility for enacting this vision resides with the Directorate of Public Works (DPW) and specifically with the Environmental Division (ENV) of the DPW. The Environmental Division of the DPW is organized into several branches, each of which is responsible for a particular aspect of environmental management on the base (Figure 1).

The Environmental Division has been a leader in the use of GIS for better management of military lands. The first implementation of the Geographic Resources Analysis Support System (GRASS) occurred at the installation in 1984. The ENV has since adopted the Environmental Systems Research Institute (ESRI) suite of products as its GIS of choice. There are hundreds of geospatial and supporting data sets depicting natural and social environments present at the base. The offices that use this data will be the primary users of the DER. Each of the main groups is described below in light of how GIS is used for resource management.

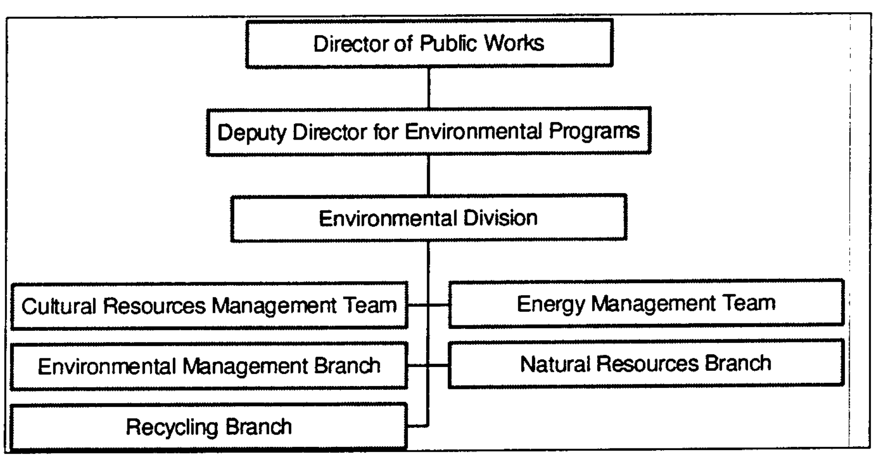


Figure 1. DPW Organization.

The Cultural Resources Management Team (CRMT) monitors and documents all cultural resources found on the installation, including both historic and prehistoric archaeological resources, cemeteries, and historic structures. CRMT further works with military trainers, the Texas State Historic Preservation Office, and the Advisory Council on Historic Preservation to prevent and ensure against negative impacts on cultural resources located at the base. A key element in the success of the CRMT is their use of a GIS database to not only store information, but also to provide spatial and site condition information used to support military trainers and management reviewers in the decision process for planning and conducting military maneuvers.

The Natural Resources Branch (NRB) is responsible for natural resources including wildlife, habitat, land management, and endangered species. Part of this responsibility includes developing and monitoring both management and long-range plans for the protection and conservation of natural resources. To assist with this responsibility, the NRB has contracted with The Nature Conservancy (TNC) of Texas, a nonprofit, wildlife conservation organization that carries out wildlife and habitat monitoring, as well as research on the base to aid in protection and conservation.

The NRB further coordinates plans and programs associated with the Integrated Training Area Management (ITAM) program. The purpose of ITAM is to provide full GIS-based support for the planning, reviewing, and conducting of military training exercises, incorporating such factors as terrain analysis, natural and cultural resources, utilities, and man-made infrastructure into the process. ITAM's mission is highly focused on geospatial technologies. Although it is not a part of the ENV, ITAM works very closely with all of the above units and plays a pivotal role in the day-to-day GIS activities at the installation.

The Environmental Management Branch (EMB) is responsible for environmental resources and managing environmental programs. As part of this responsibility, EMB provides: (1) expertise in project development and design to minimize impact on environmental functions, (2) technical training to civilian and military groups regarding environmental activities, and (3) technical guidance on environmental restoration projects (Directorate of Public Works 2000). The EMB does not currently play a central role in GIS activities at the installation, but its mission is conducive to making use of such a system.

Two other parts of ENV are the Energy Management Team (EMT) and the Recycling Branch (RB). The EMT oversees the Army's energy program and acts in a similar capacity to the EMB except that its focus resides on matters relating to

energy conservation. The RB oversees all things related to the recycling program. Neither of these two units currently use GIS.

The Engineering Plans and Services (EPS) of the DPW works with the Environmental Division to enact the installation's environmental vision by providing expertise relating to the base's utilities and infrastructure. EPS, and specifically the Support Team of the EPS, uses both Computer-Aided Design (CAD) and GIS to develop a database to provide support and service for designing, developing, and conducting construction projects, as well as providing cantonment maps.

The diversity of the natural environment, the dynamic nature of the institutional setting, and the important role the base plays in preparing for the nation's defense make it a fruitful candidate for advanced data management technologies. Fort Hood is challenging because it is not a tabula rosa, where a new system can be placed on an open playing field. It is a mature site, with a well-developed infrastructure. It is also a site where the legacy of the personal computer and traditional GIS data management strategies have led to a somewhat disjointed data landscape. The role of the DER is to build on the existing GIS activities, while making optimal use of new technologies and options to extend and improve on practices currently in place.

Approach

This report is a result of a 1-year effort to evaluate emerging technologies for sharing geospatial data and to develop a plan for the implementation of an enterprise system at a military installation. The following steps were taken to accomplish this.

1. Review of technology. This included discussions with software vendors, evaluation of beta software, and review of literature.
2. Discussion with military installation personnel. Conducted a series of meetings and communication with ITAM, natural and cultural resource personnel at Fort Hood, TX.
3. Implementation of prototype system. The prototype system was implemented at Pacific Meridian Resources in order to test its use.

2 Current Geospatial Data Management

Introduction

This chapter of the DER prototype study report addresses institutional and data management issues. The difficulties in sharing data are closely related to the manner in which data are used and managed. Technology provides the vessel for better data management, but people decide how to fill it and use it. This chapter discusses DER stakeholders, provides a description of the installation's data inventory and an assessment of accountability for specific data layers, and offers recommendations for the future.

Three conditions or objectives guide this analysis and the recommendations that follow.

1. **Corporate Knowledge**. The staff that uses data daily is generally well-informed about the location and content of geospatial data at the installation. Some of that knowledge could be shared better with a more standardized method of documentation and data access. The current method for data sharing requires tapping into the knowledge of a set of key individuals. Responding to requests for data can be an impediment to fulfilling other critical functions. The method does not function when personnel are absent and breaks down with turnover in staff. A more automated, coherent method of data access would alleviate the pressure on individuals by standardizing corporate knowledge and would provide a more efficient and effective way to share data.
2. **Accountability**. Accountability for key data layers needs to be established and documented to avoid the occurrence of numerous versions of data with the same title.
3. **Coherence**. The multiple complexities of data volume, GIS data formats, multiple shared servers, and physical distance between data users have been impediments to seamless sharing of geospatial information and related databases. All shared data needs to have a coherent "front end" so that the user sees a common, well-documented data base.

With these needs in mind, the sections that follow describe the stakeholders, current state of the data, accountability for data layers, and security issues at the installation.

Stakeholders

The term "stakeholder" has become popular in recent years. It came originally from the analysis of corporations and their social responsibility. It was coined as a play on the word "stockholder." Traditionally, corporations are accountable primarily to stockholders — those with a well-defined financial stake in the company. The stakeholder view described by Freeman (1984) approached the question of corporate responsibility from a broader perspective. Not only must responsibility extend to those with financial stake — the system will only work for the long term if all people with a stake in the outcome of an action are taken into account.

For the DER, the stakeholders include those who:

- can influence the DER in some way,
- can benefit from the DER,
- have invested resources in the system,
- have an interest in the outcome, and
- have other programs that may depend on the effectiveness of the DER.

Given this set of reasons to include a person as a stakeholder, the following types of stakeholders were identified. The first set of stakeholders is users of the system. The second set includes other stakeholders who would not use the system, but have some other interest in it.

Users

Primary Users. The primary users of the system are those who have the most interest and most invested in the system. Within this group, there are some people who will use the system as a daily part of their work and some who will rely upon the technical services of others.

1. The Environmental Division, the NRB, and the CRMT form one part of the core group of primary users.
2. The ITAM Program Office is a second part of the core users.
3. TNC of Texas operates at the base through a contract with the NRB of ENV. TNC is an important player, but operates under ENV.

Secondary Users. The secondary users are those people who currently rely on the people in the offices of the primary user group to search for and gain access to maps and geospatial data. These include:

1. The G-3 Directorate including Range Control, Range Safety Office, and Range Engineering and Planning.
2. The Engineer Research and Development Center (ERDC) personnel who are working on research and contractual projects at or for the installation.
3. Contractors, both public and private, who need access to geospatial information in order to complete their work.
4. Other governmental and administrative units with interest and/or jurisdiction in the area.

Others

This group has influence on and may invest resources in the DER, but is not likely to be a user.

1. Forces Command.
2. Directorate of Information Management.
3. Management and command level personnel.
4. ERDC LMS program.

Data Inventory

Geospatial data is used for a variety of purposes, including military training, cantonment construction, environmental monitoring, and in the planning and management of land use on the base. Currently, geospatial data is dispersed across seven different personal computers or servers located on the base. The current state of data management inhibits access to up-to-date data themes, reduces time efficiency in accessing and acquiring relevant data, and requires time of key personnel to respond to data requests.

The inventory assessed three servers and four workstations storing geospatial data in the offices of the primary user group. Two of the servers and all four of the individual computers are located within the DPW. The remaining server is located in the ITAM office, associated with the Range Safety Office and Range Engineering Office. Of the four workstations located within the DPW, two are managed by TNC of Texas, which is contracted with the NRB of the DPW. The

remaining two servers and two computers located within the DPW are managed by the Environmental Division and the Cultural Resources Division.

Table 1 provides the, physical location and description of each of the servers and workstations.

Table 1. The servers and workstations included in the survey.

Location Code	Description	Location
ITAM-1	ITAM	Bldg. 56000, #137
DPW-1	Archeological resources bldg	Building 4219 77th Street
DPW-2	Archeological resources bldg	Building 4219 77th Street
DPW-3	Directorate of Public Works Room #4	Building 4219 77th Street
DPW-4	Directorate of Public Works Room #1	Building 4220 78th Street
TNC-1	The Nature Conservancy Data Manager's computer	Bldg 1939 Rod and Gun Club loop
TNC-2	The Nature Conservancy GIS Analyst's computer	Bldg 1939 Rod and Gun Club loop

Data Access Scenario Under Current Data Management

To provide some insight into current data access conditions within the GIS community at the installation, and to further demonstrate the need and benefit of instituting the DER, this section presents a hypothetical scenario of what happens when an individual attempts to access specific types of data. The scenario deals with an individual who wants to access data relating to boundaries.

If an individual wanted to access data on boundaries available at the installation, s/he would be presented with 25 separate data themes covering such categories as Fort Hood Boundaries, West Fort Hood Boundaries, Texas boundaries, and surrounding area boundaries. Further, the different boundary data themes provide varying information such as Army Corps land boundaries, Buffer Zones, land zone designation boundaries, Nature Conservancy lands, area maps, DLG (digital line graph — roads, streams, rivers, etc.) information of Texas, etc. This data, in turn, is scattered across three different servers and one workstation (ITAM-1, DPW-2, DPW-3, TNC-2) that are located in three different offices. Although all of the data themes are in an ESRI format, the publication date and the scale/resolution is not available for any of these data themes, and only about half of the themes have a known time period for the data. What becomes very difficult in accessing this data is the lack of simplified names attached to the

data, which can provide some insight into the nature of the data. Currently, for example, if an individual wanted just a simple boundary map of the installation, they would have the following data themes to choose from:

Simplified Name	Data Location	Path Name
Boundary	ITAM-1	X:/GIS/Arcview_Data/Boundary&Training/Boundary
Ft. Hood Map	DPW-2	G:/Env/CRM/Arcdata/CRMbackup/ArcData/Ft_hood.area
Detailed Ft. Hood Map	DPW-2	G:/Env/CRM/Arcdata/GeneralData/rTas_full
Boundary of Ft. Hood	DPW-2	G:/Env/CRM/Arcdata/GeneralData/boundary_line

At present, the user would have to rely on: first, knowing where data themes for the installation's boundaries might be physically located; second, being able to access that physical location; and third, being able to locate the data themes by their data paths and saved file names. Of course, if the user wanted to know the installation's boundaries in relation to the state boundary of Texas, the user would then have to access additional data themes for this information, and there are currently seven different data themes, located across four different physical locations, which might provide this information:

Simplified Name	Data Location	Path Name
Fort Hood and Surrounding Area Maps	ITAM-1	X:/GIS/Arcview_Data/Area_Maps
Texas Maps	ITAM-1	X:/GIS/Arcview_Data/Texas_Maps
Area Maps (counties, TX, Ft. Hood boundary)	DPW-3	J:/Permanent/Area_Maps/
Various Boundary Maps	DPW-3	J:/Permanent/Boundary/
Texas Map (various data)	DPW-4	J:/Permanent/Texas_Maps/
Texas Map (various data)	TNC-2	D:/GisData/Texas/
Texas Map (various data)	TNC-2	D:/Texas/

As the above hypothetical scenario illustrates, current data management results in a potentially cumbersome situation when it comes to finding, accessing, and sharing data among the various offices using geospatial data and it adds to the data search for people off-site who require data for performing work at the installation. The DER would greatly improve this situation by not only storing all geospatial and supporting data on one central server accessible by all, but also by organizing the data and attaching simplified names and metadata, all of which should make the use of geospatial data more time and effort efficient.

Inventory Results

One part of the DER involved an inventory of all available geospatial and supporting data located at the installation. A survey was conducted from October 1999 to May 2000. To conduct this survey, personnel in the primary user offices provided information and access to computers to enable documentation of exist-

ing data. A data survey form was developed to assist in this survey. The complete set of survey information collected is found in Table 2. The intention of the survey was to collect information in a systematic manner across the servers and workstations of the primary DER users.

The unit of measure – or what constitutes the content of a survey sheet – was determined, ultimately, by the person performing the survey. It was not the intention to survey every file found at the installation, but enough detail was needed so that the DER could be populated and organized around the way information is already being used and created. The general guideline used by the survey was that each "data theme" on a given workstation or server needed to be documented as a sheet. A data theme is defined as a grouping of one or more files used to represent a singular theme. As such, many data files may be used to constitute a single data theme. Two different versions of the same theme, or the same theme for different years, did not warrant a new sheet. Groups of related files could also be entered as one theme. For example, an aviation data theme might include individual data files for all of the separate airfields located on and around the base, as well as routes and corridors. The actual file names associated with the theme were recorded on the sheet to help link the data theme with specific files on a specific computer.

Table 2. Content of data survey sheet.

Topic	Details
Simple Name	A unique but simple name for the theme
Purpose	Why created
Creator	Who created the data
Computer	Server or workstation where data are located
Publisher	Fort Hood POC
Publish Date	When made available
Date	Time period of the data
Collection interval	How often collected One time effort Regular basis Intermittently
File Format	a. Shape file/Arc coverage, Excel, etc. b. Point/line/polygon/grid/image/TIN
Spatial Reference	Coordinate system, datum
Coverage	Geographic Extent. Ends at boundaries of Fort Hood Subsection of Fort Hood Larger than Fort Hood
Security	Any concerns about access?
Requesters	Who might ask to use this data?
Quality	Any known quality issues
Files/Notes	List all the files in this theme. Provide notes as needed.

The process resulted in a set of 250 sheets. These data forms were then tallied and entered into an Excel spreadsheet from which the summary charts in this report were derived. The remainder of this section provides a summary of the findings of this inventory outlining the various aspects of the existing geospatial data.

Location of Data

Figure 2 shows the number of data themes located on each server/computer. DPW-2 houses the largest number of data themes. This server contains data primarily associated with archaeological resources. Table 3 denotes the primary, secondary, and tertiary category of data hosted by each server/computer.

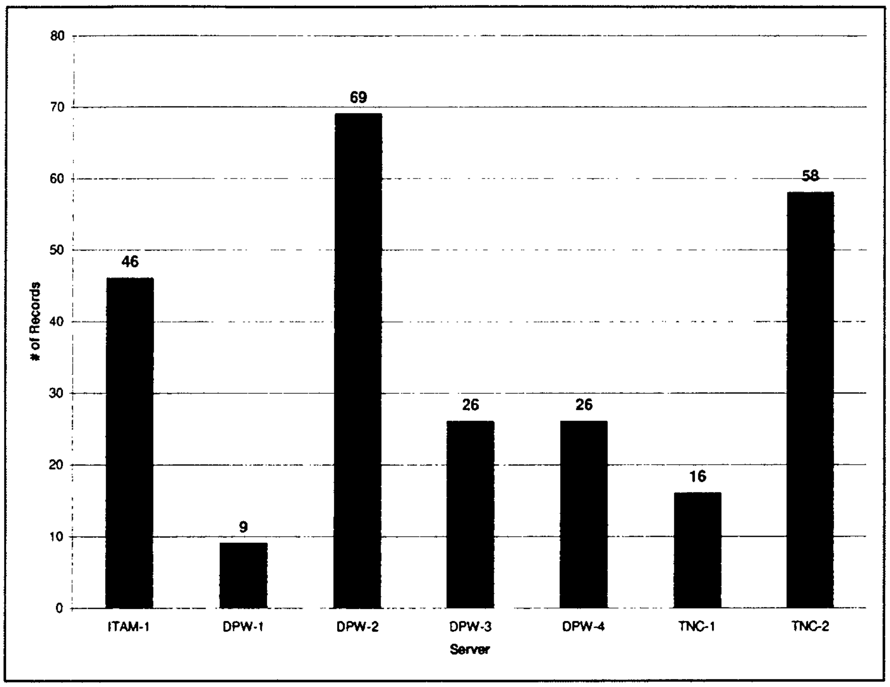


Figure 2. Number of data themes by server.

Table 3. Data theme type by server.

Location	Primary	Secondary	Tertiary
ITAM-1	Terrain Analysis	Transportation	Imagery
DPW-1	Archaeological		
DPW-2	Archaeological	Transportation	Cadastre
DPW-3	Boundary	Grids	Buildings
DPW-4	Hydrology	Transportation	Utilities
TNC-1	Endangered Species		
TNC-2	Endangered Species	Boundary	Imagery

Data Format and Compatibility Issues

The format of data is an important consideration in the feasibility of the DER. Given the state of GIS technology, many GIS software platforms are proprietary of their data formats; the platforms are therefore unable to effectively and seamlessly import foreign formats. This becomes a potential obstacle in the sharing of data and/or the flexibility to move data between varying platforms without the data becoming distorted.

As it turns out, data formats are not an issue at Fort Hood, since the GIS community at the installation primarily uses ESRI products. This significantly aids in mitigating the obstacle of data translation. The issue of interoperability among types is limited to a few distinct situations. The data inventory did reveal the existence of several different data formats, shown in Figure 3.

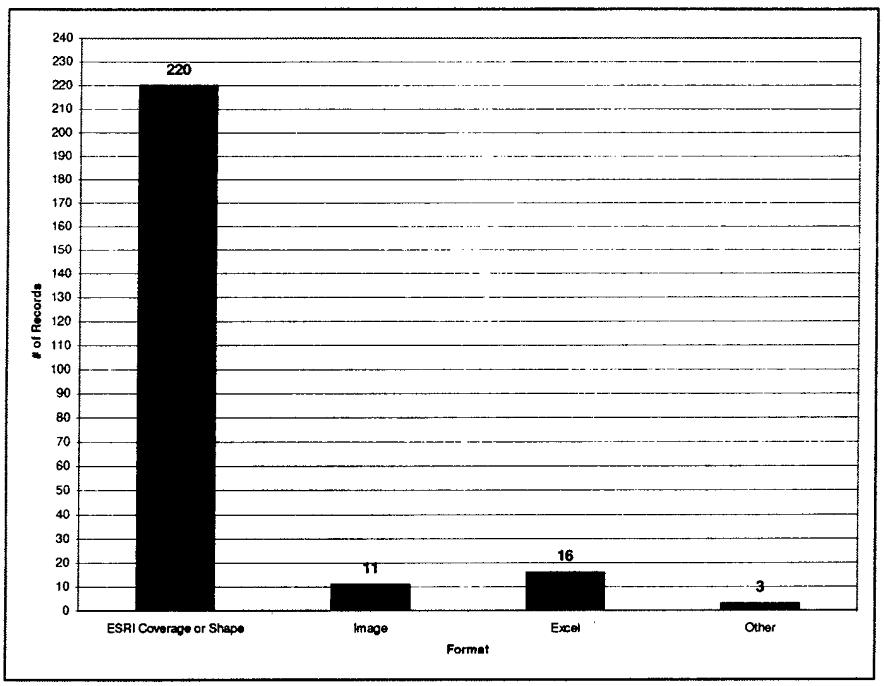


Figure 3. Data formats.

The majority of the data themes are formatted as ESRI coverages or shapes. Due to the use of ArcSDE (spatial database engine), these formats are easily accommodated by the DER. The files in Excel will be linked to the DER through Oracle, but will not reside in the Oracle database. The Microsoft Excel format is a de facto standard for spreadsheets. The eleven data themes formatted as images include satellite imagery, digital raster graphics (DRG), digital orthographic quarter quads (DOQQ), and aerial photographs. Like the Excel files, these will be linked to the DER. The next version of ArcSDE will more fully integrate image files. In the meantime, the image formats will be linked through

Oracle. The three data themes grouped in the format category of "Other" include TIN (triangulated irregular network), Microsoft Access database, and grid based DEM (Digital Elevation Model) formats. The TIN and the DEM are actually both in ESRI formats. With the proper ESRI software, they are accessible.

Several data format issues remain. The primary problems are related to the CAD drawings and their integration into the ESRI-based GIS database. The CAD files are created with Bentley MicroStation software. The CAD drawings of the installation use a non-standard coordinate system, which makes it difficult to rectify the information with the other land-based GIS data. In addition, the CAD representation of objects is geared toward a visual depiction. Text is mixed with graphics, some street segments are not attributed, and points are depicted as small circles. Each of these issues presents difficulty and requires editing to make the CAD layers suitable for inclusion in the DER.

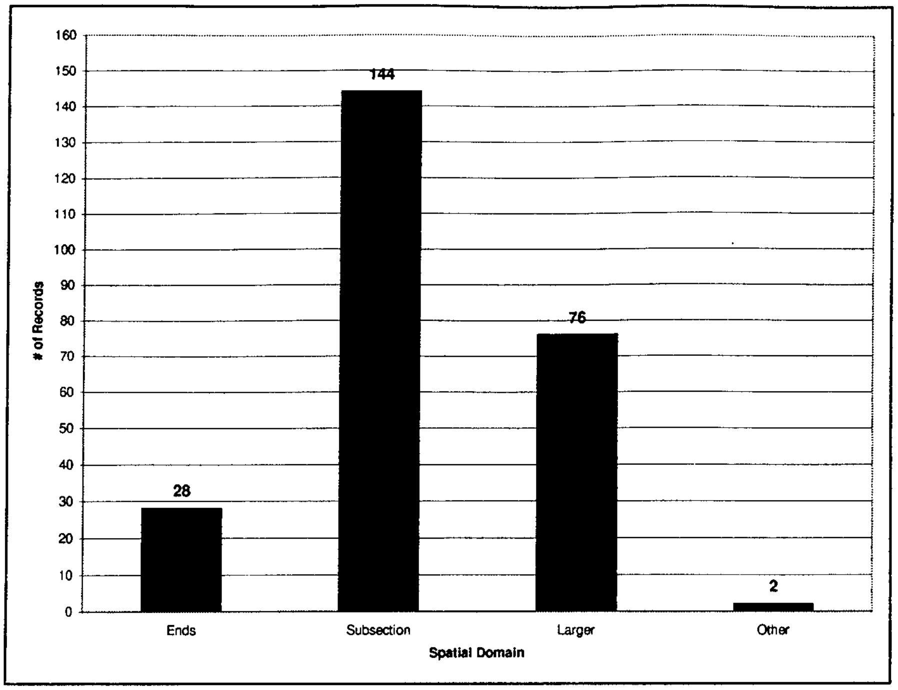
The DER cannot solve these problems, but the introduction of Safe Software, Inc.'s Feature Manipulation Engine (FME) may help to streamline the process. FME is a geospatial data universal translator that ERDC/CERL, in conjunction with a project contractor, Pacific Meridian Resources, has found to be most effective for translating data across various formats. FME is currently the most sophisticated program available for data translation among various vector formats.

A smaller set of problems is related to the use of the GRASS GIS. Although some GRASS files remain in use at the installation, virtually all of the GRASS geospatial information of interest has already been translated into an ESRI format.

Coverage and Spatial Reference

The coverage of the data refers to the geographic area of the data. The spatial reference refers to the datum, projection, and cartographic measuring unit used to represent this information. Each individual data theme shares compatible coverage and spatial reference among its composite data layers. Coverage and spatial reference can have a direct bearing on the efficient and seamless sharing of data in so far as different coverages may use varying scales and projections, which may or may not be compatible with each other when data themes are brought into a GIS platform. This issue can be resolved, in most cases, through the use of FME, which is capable of reprojecting geospatial data.

The data inventory revealed the existence of four categories of spatial coverage and five categories of spatial reference. Figures 4 and 5 reveal the breakdown of data themes based first on coverage and second on spatial reference.



Spatial Domain	Definition
Ends	Coverage is the size of Fort Hood—Domain ends at the boundary of Fort Hood
Subsection	Coverage is delineated areas within the boundary of Fort Hood
Larger	Coverage extends beyond the boundary of Fort Hood
Other	Unknown coverage--no information provided for data themes in this category

Figure 4. Spatial coverage of Fort Hood data.

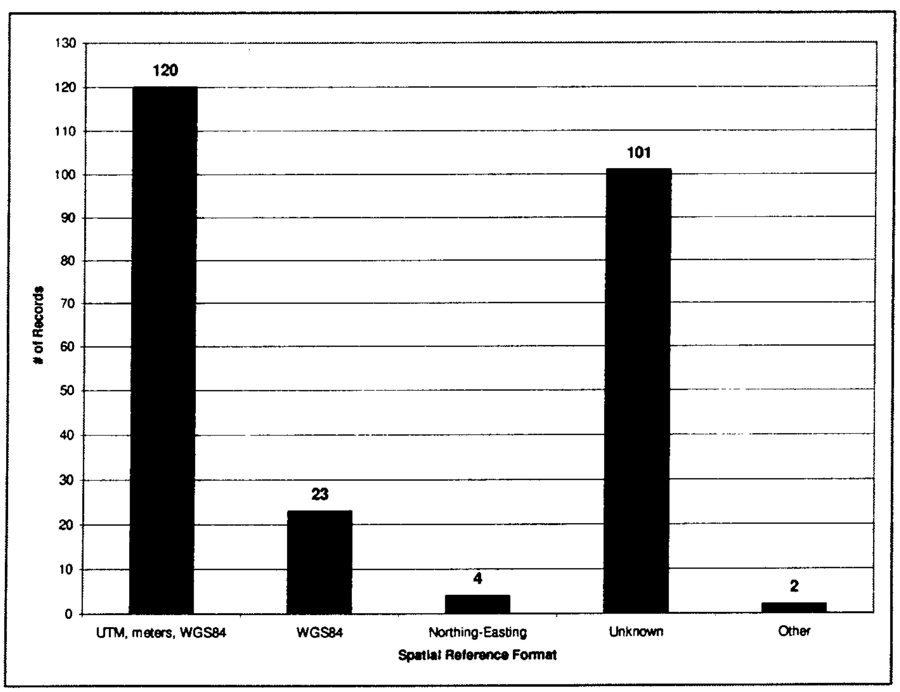


Figure 5. Spatial reference of Fort Hood data.

Although the majority of data themes have a known spatial reference format of UTM (Universal Transverse Mercator), meters, WGS84, there is a significant number of data themes for which the spatial reference format is unknown. This may prove an obstacle for the DER. The category "Other" contains two data themes that were Excel spreadsheets for which a reference was not required. In some cases, the UTM coordinate system was not verified by the survey taker. The default spatial reference for the DER will be WGS84, UTM meters.

SDS Entity Set and Thematic Classifications

The CAD/GIS Technology Center has produced the Spatial Data Standards and Facility Management Standards (SDS-FMS). The DER makes use of the Spatial Data Standard (SDS) Entity Sets to allow for the organization of data themes into predetermined categories based on the information that each data theme contains. Organizing the data themes residing on the various servers/computers by SDS Entity Sets further provides a quick reference guide, or overview, of the nature of the data existing in the GIS community at the installation.

There are 26 entity sets comprising the SDS; each entity set is further subcategorized into classes, and then types, which provides for additional clarification of data content. Each data theme was also assigned an SDS entity class, but the complexity of that assignment is outside the scope of this document. As part of the DER, the data in the DER will be assigned SDS entity sets and classes, which can then be used to reference the data for purposes of searching. Definitions of each of the above-represented SDS Entity Sets are provided in Appendix A. For more information on the SDS, see the CAD/GIS Center web page at: http://tsc.wes.army.mil/products/TSSDS-TSFMS/tssds/html/.

The inventoried data can be categorized into 20 of the 26 entity sets comprising the SDS (Figure 6). The six entity sets not represented are auditory, climate, future projects, olfactory, soil, and visual. [Note: data do exist that can be categorized into the SDS Entity Set of soil; however, this data is combined with other data that has been classified into the SDS Entity Set of geology.] Figure 6 shows one category denoted "other" that contains data themes for which an SDS Entity Set could not be clearly determined due to a lack of information. The three themes that comprise this category are old and will most likely be deleted.

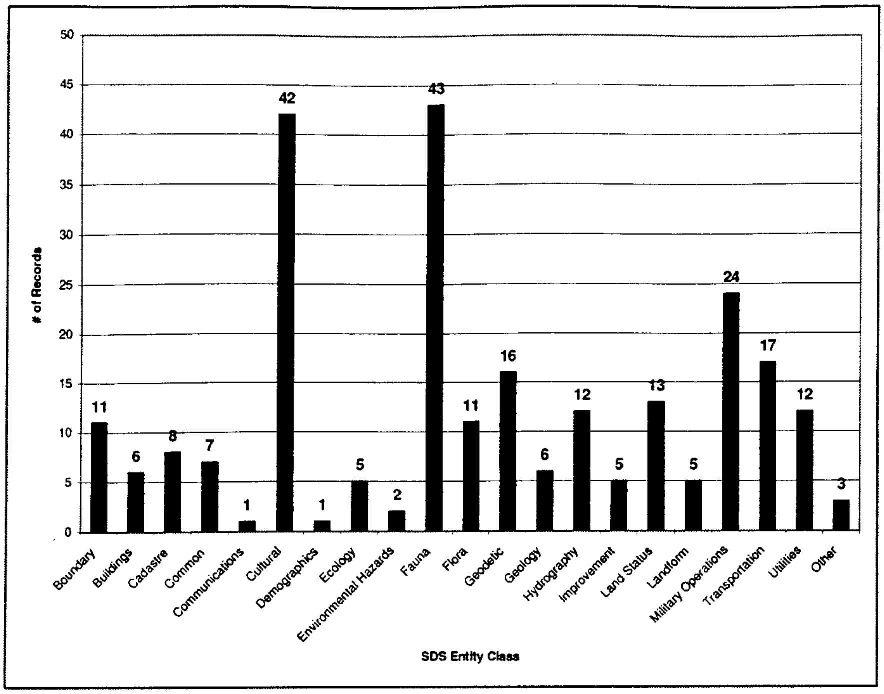


Figure 6. Spatial data standards entity set classifications.

The SDS provides a standardized nomenclature for data. The data themes used, however, are not entirely consistent with the SDS. To provide the installation a more natural view of its own data, researchers proposed the DER with a thematic organization method based on the contents of the data inventory and discussion with staff in the primary user group. This alternative organization scheme simplifies the current SDS scheme and is more tailored specifically to the data existing at the installation. This alternative organization scheme is not intended to replace SDS, but rather to complement SDS and serve as an organizational format for the user who is looking for data in the DER. Table 4 and Table 5 list the alternative scheme and its relationship to the SDS, respectively.

Table 4. Alternative data organization scheme.

General Heading	Categories
Administrative	Boundary, Demographics, Grid Scales, Land Use/Land Cover
Cultural Resources	Archaeological, Cemeteries, Management
Ecology	Endangered Species, Fire Management, Habitat, Vegetation
Human-Made Features	Infrastructure, Transportation, Utilities
Military Training	Aviation, Crossings, Land Use, Maneuvers, Safety/Danger Zones, Terrain Analysis, Training Areas
Natural Features	Caves, Elevation/Contours, Erosion, Geology, Hydrology, Recreation
Image	Aerial, DOQQ, DRG, Satellite, Other

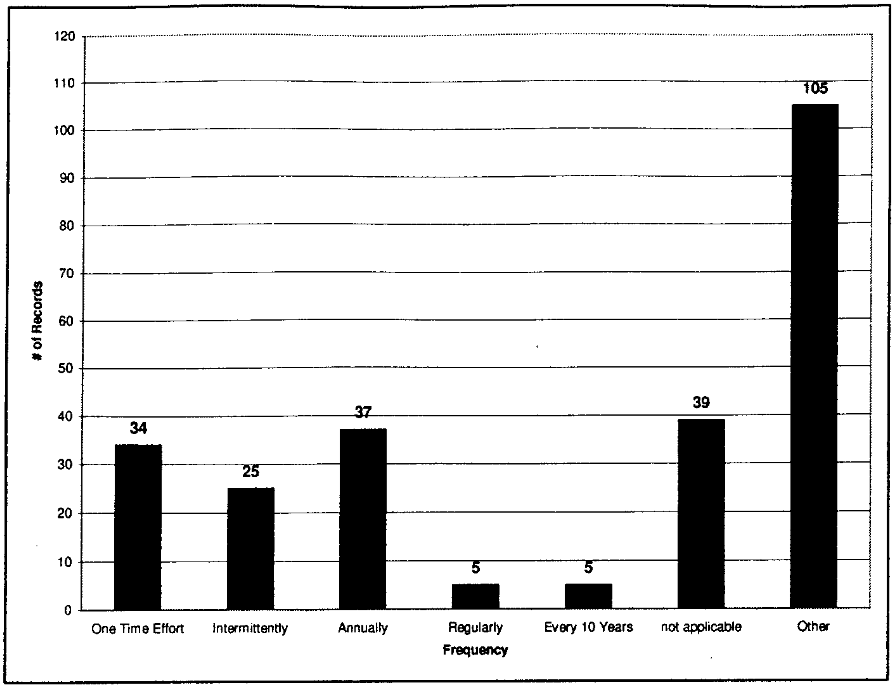
Table 5. Relationship between SDS Entity Sets and alternative data theme organizing schema.

Alternative Data Theme	SDS Entity Set
Administrative	Boundary (bd) Cadastre (cd) Common (cm) Demographics (de) Environmental hazards (eh) Land status (ls)
Cultural Resources	Cultural (cr)
Ecology	Auditory (au) Climate (cl) Ecology (ec) Fauna (fa) Flora (fl) Olfactory (ol)
Human-Made Features	Buildings (bg) Communications (co) Future Projects (fp) Improvement (im) Transportation (tr) Utilities (ut)
Military Training	Military Operations (ml)
Natural Features	Geodetic (gd)* Geology (ge) Hydrography (hy) Landform (lf) Soil (so)
Image	Visual (vs) Geodetic (gd)*

* Geodetic can be placed in more than one alternative data theme category.

Acquisition Frequency of Data

The acquisition frequency of the data existing at the installation indicates the current practices to keep data themes updated and current. The data inventory indicated that the majority of the data themes do not have a known frequency acquisition. This may indicate the presence of outdated data themes. The second largest category illustrated in the chart in Figure 7 is that of "not applicable" which refers to data that does not require being updated. While this category is not necessarily cause for concern such as with the category of "other," careful scrutiny of the data themes in this category should be done to ensure against having outdated data.



Frequency	Definition
One Time Effort	Data is collected only once.
Intermittently	Data is collected and/or updated on an irregular and non-specific schedule, or whenever changes occur or new information is added.
Annually	Data is collected and/or updated on an annual schedule.
Regularly	Data is collected and/or updated on a regular schedule at a regular frequency, other than annually, as determined by the requirements of the data.
Every 10 Years	Data is collected and/or updated every 10 years with the Census.
Not Applicable	Includes data for which acquisition frequency is not applicable whether it be because the data never changes (such as state boundary maps), or the data only has a limited purpose and lifetime, and updating or maintaining it is not required.
Other	Does not have known acquisition frequency information.

Figure 7. Update frequency.

Potential Duplication Rates

This section considers the potential for the existence of duplicate and redundant data. While determining the existence of duplicate or redundant data will require direct examination, the data inventory does provide some guidance as to where duplication and redundancy may exist on the servers/computers at the installation. Examining the data inventory indicates that roughly 49 records or 20 percent of the existing data is potentially redundant and/or duplicated (Table 6). This was determined by comparing the data across computers for duplicate or near duplicate titles. Appendix B is a listing of the data themes that comprise this statistic.

Table 6. Potential redundancy of data.

	DPW-1	DPW-2	DPW-3	DPW-4	ITAM-1	TNC-1	TNC-2
Number of Potentially Duplicate/ Redundant Records Found on Server/Computer	0	12	6	10	13	2	6
Percentage Potentially Duplicate/ Redundant Records of Total Records Located on Server/Computer	0%	17.4%	23.1%	38.5%	28.3%	12.5%	10.3%

Miscellaneous Issues Represented in the Inventory

Security Issues: Of the 250 data themes, only 11 data themes were designated as having security restrictions. Of these 11 data themes, two have "some concerns," two are for "Fort Hood Use Only," two are for "Official Use Only," four have "limited access," and one — the masked cultural resources map — is "Highly Restricted/Classified." The majority (6 of the 11) of these data themes reside on the ITAM server, and the majority (6 of the 11) of these data themes contain information on archaeological or natural resources. Table 7 lists these 11 data themes along with their reported security issues, and physical location.

Quality Issues: Only 21 data theme survey sheets reported data quality issues. Of these 21 data themes, 12 reported that there was nothing known about the data and so data quality is unsure, while four reported that the data quality was high. The remaining five data themes reported unique circumstances regarding data quality, including: (1) "Dirt landing strips updated with imagery; 2 major airfields done by engineers" (Aviation); (2) "Not 100% accurate because so many different crews had discrepancy in data collection procedures" (Geology & Soils); (3) "Old Data" (Landform and Topography); (4) "No known methodology of classification" (TM_Satellite); and (5) "From appearance, it looks as if it doesn't match up with boundary line of Ft. Hood (universe data layer); maybe poor quality base map data or something in the digitizing process" (Army Corps land).

Data Requesters/Users: Question 13 of the data inventory survey form asked for information on who typically uses or requests, or might use or request, the data. Forty-one survey forms provided answers to this question. The responses were varied, with no clear pattern of users/requesters, but can be loosely grouped as shown in Table 8. Table 8 further provides examples of the types of data themes that each group uses or requests.

Table 7. Data themes and their related security issues.

Sheet #	Simple Name	Security Issues – As reported on survey sheets	Server Code
2	Aviation	Some: authorized personnel only (FOUO); Gray converted to multiple use so under construction and no security issues	ITAM-1
10	Cultural Resources	Highly restricted; sensitive	ITAM-1
11	Digsites for training	For official use only	ITAM-1
43	Digital Elevation Model	Ft. Hood use only	ITAM-1
44	Feb 99 DOQQ	Ft. Hood use only	ITAM-1
46	Surface Danger Zone	Some concern or limited access	ITAM-1
53	Prehistoric Archaeological Sites	Some concern or limited access: exact locations of sites are secure	DPW-1
141	Archaeological sites (5152)	Limited access	DPW-3
145	Archaeological sites in golf course	Limited access	DPW-3
154	Endangered species bird habitats	Some concern	DPW-4
171	Endangered species bird habitats	Some concern	DPW-4

Table 8. Data users/requesters and data examples.

User/Requester	Examples of Data Themes Used/Requested
Military Trainers	Digsites, Training Lanes, Pipelines, Elevation, Training Areas
Military Units	Live Fire, Aviation, Endangered Species Habitat
Planners	Geology & Soils, Endangered Species Info,
Biologists/Researchers	Study Areas, BCVI Data, BCVI Point Locations
Land Managers	GCWA Point Locations, Brown-headed Cowbird Data
Emergency Units	Fire Roads, Aviation
Cultural Resources Management	Archaeological sites
Directorate of Public Works	Fire History, NCRS-Erosion Study, Hydrology
The Nature Conservancy	GCWA Habitat Model, Live Fire Area Mask, Buffer Zones
ITAM	NRCS Vegge Study, Crossings

Accountability

The primary user group identified the need to establish accountability for specific data themes. Accountability for data means that a particular office takes responsibility for the timely content, format, and availability of the data in question. Each office must work out the best means to accomplish this. In some cases, one individual will have both technical and subject area expertise and can accomplish all aspects of data creation and processing. In other cases, technical expertise beyond the staff of the accountable office will be required. In any case, it is expected that the accountable office will perform the tasks necessary to accomplish the goal of keeping the data up to date and well documented.

The general thematic categories (Table 4) were used to guide the accountability assignment. This was done through interviews with the parties with the greatest involvement in the DER. The table was further circulated for comments and the final result is found in Table 9. The role of the Nature Conservancy of Texas was an item that required some attention. The Conservancy plays a vital role on the base in monitoring wildlife. However, because they are contractors, they will not be given the responsibility of being accountable for the content and update of any data. They will perform this work only under the umbrella of the contract they have with the Natural Resources branch.

Accountable parties are defined as offices, not as individuals. The following accountable parties are included in this assessment.

1. ENV – The Environmental Division of the DPW.
2. NRB – The Natural Resource Branch of the Environmental Division, including The Nature Conservancy, which is under contract to the NR.
3. CRMT – Cultural Resources Team in the Environmental Division.
4. ENG – The Engineering Division of the DPW.
5. ITAM – The office of the Integrated Training Area Management program.
6. G-3 – Range Safety Office/Range Engineering Office.

Table 9. Data accountability for data categories.

General Heading	Categories	Sample Sub-Categories	Accountability
Administrative Boundaries not necessarily visible on the landscape			
	Boundary	Ft. Hood, West Ft. Hood, Surrounding areas/Texas	ENV
	Demographics	Census data	ENV
	Grid Scales	Grids, Tick marks, Indexes, Quadrant #s	ENV
	Land Use/Land Cover		NRB
Cultural Resources Of archaeological, historic, or cultural significance			
	Archaeological	Sites, Shelters	CRMT
	Cemeteries		CRMT
	Management	Archaeological sites, Buffers, Unnamed sites, Protected sites, Vandals, General cultural resources	CRMT
Ecology Fauna, flora, fires, and habitat			
	Endangered Species	Golden-cheeked warbler, Black-capped Vireo, Texas Horned Lizard, General	NRB
	Fire Management	History, Roads, Vegetation	NRB

	Habitat	Golden-cheeked warbler, Black-capped Vireo, Eagles, Brown-headed Cowbird, Study areas, General	NRB
	Vegetation	General, Forest	NRB
Human Made Features Infrastructure, improvements or structures built by humans, above or below ground			
	Infrastructure	Buildings, Corrals, Sidewalks, Features, Wells, Structures, misc.	ENG
	Transportation	Railroads, Roads	ENG
	Utilities	Communication, Pipelines, Power, Sewers, Water, General	ENG
Military Training Required for the planning and conducting of training on the base			
	Aviation	Location, Boundary	ITAM
	Crossings	NIMA	ITAM
	Military Land Use	Land groups, Platoon areas, Company area designations	ITAM
	Maneuvers	Brigade support areas, Digsites, Training lanes, Land descriptions	ITAM
	Safety/Danger Zones	Live fire, Ranges, Surface danger zones, High hazard areas, Off-limits areas	G-3
	Terrain Analysis	DLGs, Environment, Vegetation	ITAM
	Training Areas	Tank trails, Land designations, Training areas	G-3
Natural Features Attributes of the landscape			
	Caves		NRB/CRMT
	Elevation/Contours	DEMs, Contour lines	ENV
	Erosion		NRB
	Geology	Soils, Land descriptions, Karst features	NRB
	Hydrology	Dams, Rivers/Streams, Water bodies, Watersheds, General	NRB
	Recreation	Deer/deer hunting, Off-road vehicles, BLORA	NRB
Images Photographic or other digital images			
	Aerial photos	Digital orthophotos, Digital air photos	Accountability is in office that contracts the product.
	Scanned images	Digital raster graphics, Other scanned images	
	Satellite images		

Conclusions and Recommendations

The data management conditions at the installation are stable. An incremental approach to DER implementation would provide the least disruptive path to im-

proving data sharing and access. The direction for the next set of activities needs to be organized by the primary users. Items to be considered should include the following:

1. Determine responsibility for making certain that the accountability requirements are agreeable and executable for all involved. Individuals need to set up plans to execute this responsibility as needed.
2. Determine logistics for providing the correct data layers for input to the DER. These data must be in proper form and be accompanied by Federal Geographic Data Committee (FGDC) documentation.
3. Make a plan for the fuller adoption of the SDS for the standard data layers in the DER.
4. Define the options for a strong organizational element to provide on-going coordination of GIS activities.

Additional recommended actions include:

1. Use Feature Manipulation Engine from Safe Software, Inc, to facilitate the translation of data from the Bentley MGE format to an ESRI coverage format. This would involve setting up a custom "mapping file" that carries over attributes and spatial entities from the CAD to the GIS file formats. This is not the black box approach that is available with many GIS import/export tools. With FME, one can create a custom configuration that not only extracts specific features from the source data, but also can assign additional attributes to the data or change the spatial type of the feature extracted to a different type.
2. After determining the content and supplying the data files for loading into the DER, the accountable parties should also ensure that old versions and copies of the data are no longer present on the servers and workstations. This will help to address the data integrity issues and will wipe the slate clean for future data management.
3. Coordinate with GIS core group to ensure consistency of spatial reference system and its implementation in ArcSDE. The data in the DER will not be tiled and it will be in one spatial reference system. The spatial reference system will be WGS84 and UTM meters. ArcSDE stores data in its own coordinate space. The implications of how this is set up needs to be discussed and agreed upon before loading data into the DER. The GIS core group needs to ensure that the coordinate space in the DER is suitable for their needs.

3 Technology and Data Storage

Vendor Options and Technology Trends

The time is ripe for the sophisticated development of a centralized distributed system for geospatial data storage and access. Great strides have been made recently in server hardware and software, especially in response to the demands of e-commerce. Database Management Systems (DBMS) are more robust and better able to handle a wide variety of media types. For geospatial applications, the introduction of Standard Query Language/99 (SQL/99), the most recent standard language for DBMS (Gorman 1999), opens up many new options for storage of multi-media data and object-oriented (OO) data models. GIS development has benefited from these and other trends, as the approach to GIS data and applications has come closer to the mainstream of information technology.

An assumption was made early on in the DER evaluation process: the DER architecture would be fully compatible with and build on the existing system at the installation. The environmental GIS users on the base use the ESRI suite of software products; thus, the ESRI approach plays a central role in the assessment and recommendations.

In light of the requirement to use the ESRI software, it is important to note some of the ESRI's recent changes. The traditional flagship ESRI product, ArcInfo, has been a leading choice of a full-featured GIS, and the ArcView product is leader among desktop solutions. In the past year, ESRI has introduced a major conceptual and technical change in its approach to GIS. ESRI's primary innovation is to embrace object-oriented programming and data models with the newly defined Arc8 and ArcView8 products designed to take the place of ArcInfo and ArcView. Other vendors have products (Smallworld, for example) that have long embraced OO principles. For ESRI users, this approach is new territory and the change in approach plays an important role in the changes in data management practices at the installation.

The decision to use the ESRI approach simplified the DER software and hardware options. At the same time, a number of issues arose in the process of working out the details of the system that reveal the complex nature of the implementation of the enterprise GIS. Some of the lessons learned while determining the

DER architecture are included in the following section, which also describes the final recommended architecture. At the time of this writing, the case study installation had not implemented an enterprise GIS, and the final configuration is not known.

Architectural and Functional Overview

The system is described in two ways. First is the system overview perspective, which provides information on the major software and hardware and supporting components and how they are related to each other. For this section, the system software and hardware components are discussed. Then, the system framework is divided into the user, application, and data tiers. Second is a description of data storage issues when using ArcSDE and Oracle in the case study context.

Architecture Overview

The system has several main components, depending on the type of access needed to obtain data or maps from the system. To understand the need for the various components, it is important to describe the types of access that are available. Access to the data in the DER is provided through three separate paths.

1. Direct access to database. For core GIS users, the ESRI Arc8 software tools can be used to link directly to the data, which is stored in an Oracle database supplemented by ArcSDE. The ArcCatalog has an option to link to a database. Once the link is made, the data in the DER are simply part of the data available in the GIS environment. The data are available in either read-only or read/write form, depending on the data type and the user access level.
2. Web-based access. For non-GIS users, for users located offsite, or for those who require data to be uploaded into the system, a web-based access application is available. This method facilitates the submission of data for the DER from a variety of people. It organizes searchable fields and indexes data layer identifiers (IDs) for ArcSDE and ArcIMS (Internet Mapping System) applications, and it lets users search on the contents of the DER. The application also lets users view and print maps from the DER and possibly download data from this forum. The ability to download data would be subject to the user permissions and the sensitivity of the data. This application is described more fully later in this chapter.
3. Access for applications. The third method of access to the data is from applications. If necessary, the data can be stored in such a way that access is possible directly through tools in Oracle 8i. Alternatively, the architecture proposes the

use of the ArcSDE Application Programming Interface (API). This portion of the system has not yet been fully tested.

The prototype Data Enterprise Repository system is made up of the software components as listed in Table 10. Three separate servers hold the software for the DER (Figure 8). All of the servers at the installation are NT based. One server, the "Data Server," holds Oracle 8i RDBMS (Relational Database Management System) and ArcSDE 8.0 services. Another server, the "Web Server," is where the web application, server software, and ArcIMS reside. This server would need to be outside the firewall in order to be accessible to the Internet, and thus to people off-site. It could also be kept inside the firewall if Intranet (within the installation) use is all that is required. The final server, the "Transaction Server," may also be in the arrangement to serve Oracle and applications that involve numerous transactions. This segregates the DER geospatial components from the business uses of Oracle.

Table 10. Software components of the DER.

Component	Version	Description
Oracle 8i	8.1.6	Core relational database engine that holds tabular and spatial data.
ArcSDE - Spatial Database Engine	8.0.1	Database add-on that spatially enables the database. Core ESRI technology.
ArcIMS – Internet Mapping System	3.0	Scaleable internet mapping solution that interfaces with ArcSDE.
Arc8/ArcView8	8.0.1	Data manipulation tools from the core Arc/Info v8 release.
Windows NT server	4 Sp4+	Server operating system.
Microsoft Internet Information Server (IIS)	3 or 4	A web server that is part of Windows NT Server. Facilitates serving information to the Internet.
Microsoft Transaction Server (MTS)	Part of IIS	Part of IIS that manages object resources used by the web server and web applications.
Safe Inc.'s FME Objects – FME	2000	COM objects which can be used in custom applications that require data transformation functionality.
Custom Component Object Model (COM) components	n/a	Some functionality will be encapsulated into custom components written with Visual Basic.
Active Server pages (ASP)	2.0	A web application development platform that runs natively on Windows NT IIS.
Microsoft Access DBMS		Interacts with ASP for web application. Holds search keywords and records for SDE/IMS layers.
Microsoft Internet Explorer		Web Client for Web Access.

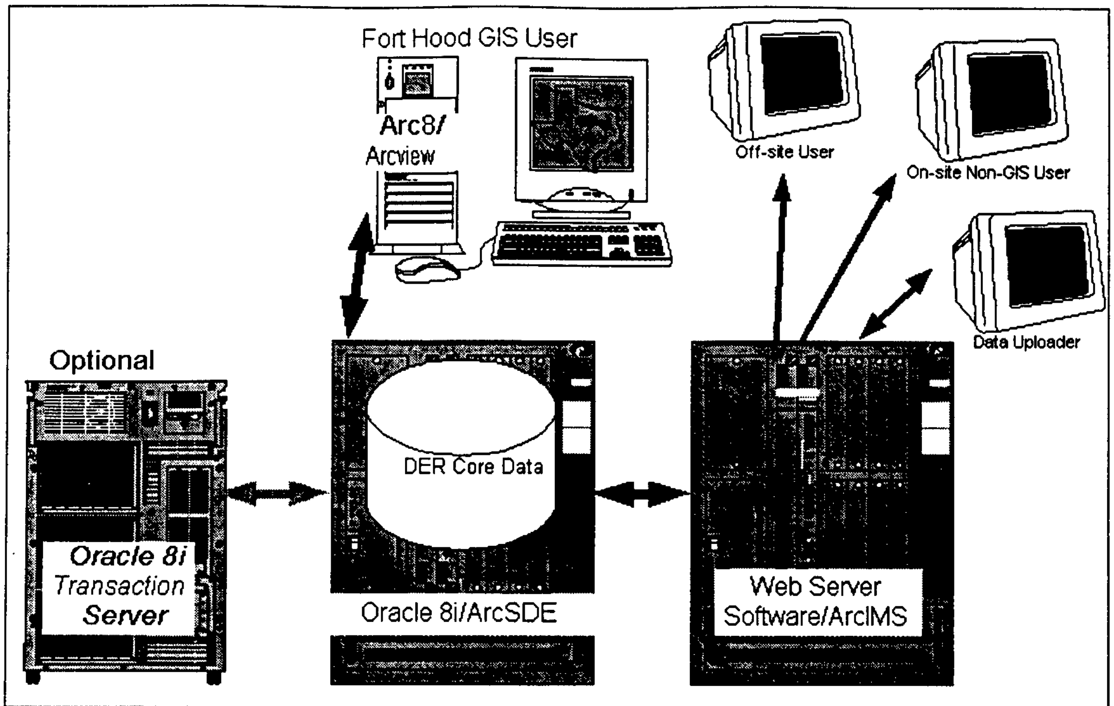


Figure 8. Simplified physical setup of the DER servers and users.

In some organizations, Oracle "belongs" to the database group and this group may not participate directly in the enterprise GIS. The ESRI recommendation is that Oracle and SDE reside on the same server. Oracle Database Administrators (DBAs) learn not to put anything on the server other than Oracle. The possible configurations to avoid this and/or evidence to quell the fears of DBAs about decreased server performance are still evolving at the time of this writing. It is possible that there is a good stable solution where Oracle is on a server separate from SDE, but it has not yet been well documented. Other servers may be involved indirectly with the DER for additional data and/or processing intensive applications. This would include modeling applications and CAD or other spatially enabled applications that need dedicated processing and data space.

Tier Technologies

Another view of the system is provided through the use of a multi-tier architecture, also known as N-tier architecture. N-tier architecture builds upon the successes of 2-tier (client-server) architectures, while removing barriers to scalability and reuse, and is easily adaptable to use across the Internet. With this architecture, the user (presentation), business, and data tiers are logically separated, as illustrated in Figure 9. To create highly scalable applications, resources such as database connections must be shared. Instead of having each client connect directly to the data resource (as in client-server architectures), client applications communicate with business services. The business services in

turn manage connections to the data resources. One instance of a business service can support multiple clients, thus reducing the required connections to the data services. These services are not required to reside on a different physical machine from the database, but they can, once again increasing scalability.

User (Presentation) Tier Technologies

- Arc8/ArcView8
- Active Server Pages
- Internet Explorer, and
- ArcIMS

Core GIS users can use the system directly through a database connection initiated by the ESRI GIS clients, Arc8 and ArcView8. Alternatively, the DER is accessed via a web browser. For the web browser, the user tier is composed of a number of Internet technologies. A framework based on Active Server Pages (ASP) comprises the main navigation and information organization. A number of ASP-based applications are present in the framework. These include a metadata search tool, a data upload toolkit, an on-line data-browser built with ArcIMS, and a data retrieval toolkit. Each of these pages uses HTML (HyperText Markup Language), ASP, JavaScript, DHTML (Dynamic HyperText Markup Language), and Style Sheets to some extent.

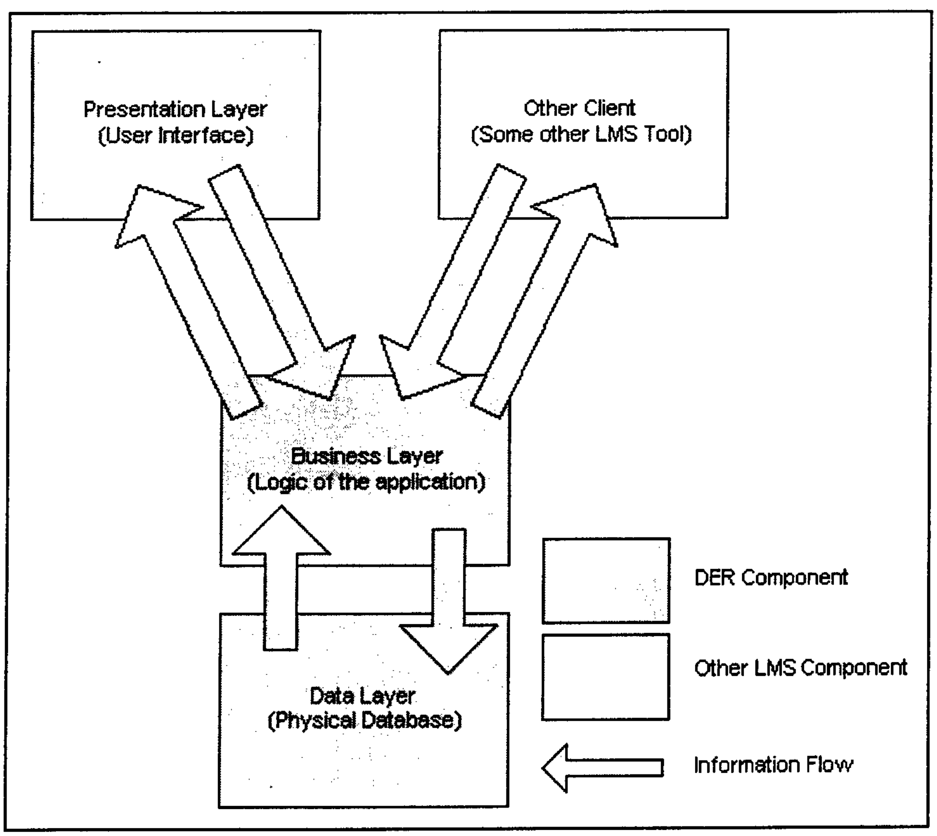


Figure 9. N-tier architecture overview.

Business Tier Technologies

- FME Objects
- Microsoft Transaction Server
- Custom Components
- Windows NT Server
- Internet Information Server (IIS), and
- ArcIMS

The business tier is composed of COM (Component Object Model)-based objects, written in Visual Basic. These objects will be accessible by not only the presentation layer, but also by other servers or clients. Accessibility is facilitated through the use of Microsoft Transaction Server, and IIS on a Windows NT Server. This solution will address the on-going DER requirements for growth and open communication between systems. Business layer tasks include data conversion into and out of the DER, data access control, and metadata input validation. To meet the requirements for loading and unloading data in a wide array of formats, Safe Inc.'s FME Objects are used for the data conversion tools. ArcIMS also has business tier components that manage the tasking of the map creation processes, as well as a site manager component for general ArcIMS administration.

Data Tier Technologies

- Access
- Oracle, and
- ArcSDE

The data tier is made up of Oracle 8i RDBMS in conjunction with ArcSDE 8.0. Microsoft Access is used to store information used by the ASP framework, while Oracle stores the spatial data. The majority of interactions with the data tier will be through ArcSDE 8.0 interface. Although Oracle is the RDBMS of choice, the prototype DER has used both Oracle and SQL Server (an alternative DBMS). ArcSDE can be used to move databases seamlessly across supported DBMSs, so the choice of Oracle was based on the preference of the installation. The ArcIMS server components will provide most of the read access to the GIS data layers stored in the spatial database.

Functional Overview

As noted earlier, the geospatial and related attribute information is stored using ESRI's SDE with Oracle. Other RDMBSs that ArcSDE supports include MS SQL Server, IBM DB2, and Informix. Oracle was chosen based on its wide use

in the Corps of Engineers, the decision of installation personnel, and because of the interest in linking to LMS applications.

ArcSDE is a middleware product that merits discussion, since it is at the center of the DER. The newest version of SDE is called ArcSDE, to indicate that it is part of the Arc8 family of software. SDE is the door to Oracle (or other DBMS) for ESRI clients. Data go through SDE when loaded into the DER, and, for the most part, data access is through SDE. Given its importance in the scheme at hand, some details are provided here that summarize the ArcSDE functions that must be considered when setting up and using ArcSDE in an enterprise solution.

Coordinate System and Projections. ArcSDE has its own coordinate system and its own means for doing projections. All spatial data in the SDE database format is stored in a coordinate system with coordinates ranging from 0 to 2147483647 (coordinates are stored as 32 bit integers – 64 bit in the future). The 0,0 point is in the lower left corner. So if you have a set of data stored in decimal degrees (DD, e.g., -80.34689,45.37777), you must provide SDE some information to translate the coordinates from decimal degrees into SDE coordinates. These pieces are:

- Scale Factor: You must supply a scale factor to account for the precision you want to preserve when the DD data are read into the SDE integer system. For example, if you want to keep five places to the right of the decimal place, then the scale factor would be at least 10,000. In addition, you need to decide how much of the SDE coordinate space you want to fill up with the data. The SDE coordinates snap to coordinate locations. You must find the balance between taking up lots of extra space versus losing detail in the original data.
- X and Y Offset: This tells SDE where in its coordinate system to locate the data being loaded. You want to make sure that all of the data in the original dataset fits into a valid part of the coordinate system – there are no negative coordinates in SDE. For example, in the DD situation, if you used the 0,0 point in the data (located at the intersection of the equator and prime meridian), three of the four quadrants would fall outside of the SDE coordinate space. A second example of the need for the X and Y offset is where you have a set of individual files that represent tiles of data. You can store these in SDE as a seamless data set, but you have to leave space for all of the individual tiles.
- Projection: The proj.txt file on the system contains parameters about a variety of projections. You can choose one of those projections by code when you load the SDE layer. That code is stored along with the data and is read by the client for display as needed.

Based on current practices at the installation, the geospatial data will be stored in the DER using UTM coordinate system in meters, Zone 14, and the WGS84 geodetic system. Most of the data in the DER is for the installation or a subsection thereof. The ArcSDE coordinate location needs to be determined through discussion with installation GIS users to preserve the proper amount of precision in the SDE layers without using too much disk space.

ArcSDE Geometry Data Structure. ArcSDE provides for the creation of several different formats for geometry storage (*Understanding ArcSDE* 1999; see also ArcSDE CD documentation for more information). For Oracle implementations, three formats are available:

1. SDE BINARY: readable only by SDE and SDEAPI applications. Stored as a Binary Large Object (BLOB).
2. Oracle 8i Spatial: readable through SDE, SDEAPI, and Oracle Spatial Applications.
3. Normalized Geometry: readable through SDE, SDEAPI, and OpenGIS compliant tools. This was known as Spatial Cartridge in versions of Oracle before 8.1.5.

The ArcSDE documentation notes that there is increasing overhead with each of these options, at least from the ArcSDE perspective (which is all ESRI is concerned with). That means that the most efficient way to store and access the data within the ESRI framework is to store it in SDE Binary, the ESRI native format. The most open, but least efficient data storage option is the Normalized option. The use of Oracle 8i Spatial opens up the possibility of better integration with systems that support Oracle Spatial.

For the current version of the DER, the SDE Binary geometry is recommended. There was no requirement to link directly to other systems, because the translation of CAD to GIS formats is the favored method of integration. The full functionality of the ESRI object-oriented data model (described below) is available only through SDE. The behaviors associated with a data object would not be available to a non-ESRI client reading the Oracle Spatial formats, for example. As the system matures, there will likely be reason to use one of the other options.

Geodatabases and Data Models. With the advent of Arc8, ESRI has introduced the concept of the geodatabase (see MacDonald 1999). A geodatabase is:

> [a]n object-oriented geographic database that provides services for managing geographic data. These services include validation rules, relation-

ships, and topological associates. A geodatabase contains feature datasets and is hosted inside of a relational database management system (*Campus Glossary* 2000).

The data model used by GIS software in the past has been file-based, and the spatial information has been segregated from the attribute data in a vendor-specific solution that involves a set of files that each serves a different function. These functions include files for holding geometry, attributes, topology, indexing, and other types of information, depending on the spatial format. The geodatabase makes several important changes to this model.

With ArcSDE, the information that was formerly in separate files is now arranged as tables in the DBMS. At minimum, three tables are needed:

- Business Table: stores attributes for the layer. It has the same name as the layer (e.g., roads, boundary). The attributes table holds information like road type, condition, etc. Each row in the table is a single feature. It also has a column with a unique feature ID that links the attribute information with the spatial information. The client talks directly to the business table. The business table is linked to the feature and spatial index tables described below.
- Feature Table: stores the geometry. In SDE binary format, the spatial data are held in binary large objects (blobs) in a field in the feature table. Raster data may be stored as a blob or the field might contain a pointer to a location on the file system where the raster data are stored. Some other standard fields in this table are entity type, envelope (the smallest rectangle that fully contains the feature), and area, length, and number of coordinates in the feature.
- Spatial Index Table: indicates in which spatial index grids a feature is located. A spatial index is a grid defined by the user when a data layer is loaded. It is similar to the ABC-123 grid commonly found on maps. It decreases the time needed to locate a feature, because you do not have to scan the entire map if you know what grid it is in. For example, if Green Street is found in D7 and D8, you can confine your search to just those two cells. For SDE, the spatial index is built after the layer is loaded in. Any spatial search request starts with the index.

With the ArcSDE data model, all data can be ensured of having a common spatial reference and tiling is eliminated. The spatial index speeds up access to large datasets since data is delivered only for the area of interest. Applications that need to get data from the DER know where to find data and what it con-

tains. Programmers who have built applications on file-based systems know what an advantage this is.

Geodatabases can also include other aspects of the data model that provide greater "intelligence" or "behavior" to the data. These behaviors include attribute domains rules as well as rules for splitting or merging features along with the ability to define subtypes. These behaviors are described briefly below. The somewhat confusing thing about the SDE geodatabase is that while it **may** include one or more of these behaviors, it is not **required** to include any of them. Geodatabases vary vastly in complexity. The simplest way to store data in the geodatabase is as independent feature classes (also called feature layers). If a coverage or shape file is imported to a geodatabase, and nothing else is done, then it is really just the same data as before but in a different format. The geometry and attributes are stored and accessed differently, but that would be transparent to the user.

Another aspect of geodatabases that can be confusing is that while all data in SDE is stored in a geodatabase, it is also possible to CREATE and use a "personal" geodatabase without SDE, using ArcCatalog. The only way to store data for an enterprise solution is through the SDE, but the geodatabase functions are available without SDE by means of the personal geodatabase. With a personal geodatabase, the tables are stored in MS Access format.

So on the one hand, the geodatabase can be used to simply store the same set of data layers that already exist. On the other hand, the potential is available to change radically the way in which geospatial data are modeled (Zeiler 1999). Feature datasets are the foundation for a geodatabase data model. Feature datasets are coherent collections of feature classes. They are designed such that a change in one layer in the dataset automatically changes appropriate data in another layer. Imagine a feature dataset of an installation's boundaries; if the boundary were to change, the training area adjacent to the boundary could be updated at the same time as the boundary layer.

With the geodatabase, the user has the ability to define the allowable "domain" of a particular attribute. To illustrate this, consider the attribute of training area name for the layer of training areas. If, in the "real" world, all of the training area names are numbers between 1 and 65, then the domain of training areas is the real numbers from 1 to 65. After this domain rule is set up, the system would not allow a user to enter 66, 0, or 106A, for a training area name. Other types of rules could be employed that involve relationships between classes. If excavation is not allowed in training areas that contain nesting birds of particu-

lar species, then an attempt to enter a digging permit request for that training area would not be allowed.

In addition to the site-specific examples provided above, the new geodatabase data model can be used to deploy standard, well documented, robust, object-oriented data models for important features shared by many land managers. The ArcGIS Hydro Data Model (GIS Hydro 2000 Pre-Conference Seminar 2000) is one example. The Hydro Data Model is documented with the Unified Modeling Language, "...a graphical language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system" (Booch, Rumbaugh, and Jacobson 1999; see also Fowler and Scott 1999). This means that it uses a standardized manner in which to describe the data attributes and behavior. In addition to the document, the ArcGIS Hydro Data Model is delivered in an Access table according to ESRI's prescribed manner. This data can be read into ArcCatalog and then "filled" with place-specific data. The model as it stands strives to "be an *essential* data model, not an *exhaustive* data model" (GIS Hydro 2000 Pre-Conference Seminar 2000). Further, it attempts to synthesize the concepts of hydrography (the way water is portrayed on maps) and hydrology (the study of the movement of water). As such it is useful to both create maps of rivers, lakes, and canals and to model the behavior of those bodies. The model is customizable so the user can adjust it to match specific needs, using the same set of geodatabase options illustrated above.

For the installation, the quickest way to implement the DER is to load data using ArcSDE or ArcCatalog directly from the existing coverages and shape files in use. This process would create a set of feature classes (layers). The procedure for doing this is described more fully in the next section on personnel. Selection of which layers to include was also discussed in connection with accountability for various layers in Chapter 2. After loading has been completed, it would be worthwhile to provide training to GIS personnel at the installation to enable them to start considering how the data requirements could be met through enhancements available through the geodatabase. The benefits of a more robust data set would help to justify the high cost of ArcSDE, which would be underused if it is used only for replicating the current data.

As a final note about the geodatabase, to fully understand and exploit the potential of the object-oriented data model, training in this area needs to include three separate thrusts:

1. ArcSDE - data loaded into SDE format are, by default, in a geodatabase.
2. ArcObjects - used to add custom behavior to data.
3. ESRI specific advice about geodatabases.

Data Storage With SDE and Oracle

SDE Loading Candidacy

In building the prototype DER, a set of sample data were loaded into ArcSDE. Most data was translated into an ESRI format and then loaded. To deal with managing file formats that are not easily loaded into ArcSDE 8.0 (mainly raster imagery, EXCEL files, and documents), a simple linked data storage system was developed (Figure 10).

Data_Layer		
PK	**DataID**	**N-Auto Counter(10)**
	Item_Name	C-Fixed Length(255)
	Layer_Name	C-Fixed Length(255)
	Load_Date	T-Date & Time
FK1	User_ID	N-Signed Integer
	Internal_Storage	L-True or False
	Link_Path	C-Fixed Length(255)
	Source_File_Path	C-Fixed Length(255)
	MetaData_Path	C-Fixed Length(255)

Figure 10. Data layers.

Data Loading Procedures and Notes

The sample data used to create the prototype system are discussed later along with the storage option for each. All the ESRI formatted data (coverages, shape files, and E00 export files) load into ArcSDE 8.0 smoothly. Free-formatted files (Excel files, for example) can be loaded into an RDBMS only if they fit very strict content and organizational rules. As a broad file type, it is best to link to the Excel files.

Non-ESRI formatted data posed more of a challenge to load into ArcSDE 8.0. It was anticipated that SAFE Inc.'s FME Objects would be used for conversion from nonnative ESRI formats into coverages, or shape files, which would then be loaded via ArcSDE 8.0's command line tools. The FME toolkit facilitates direct translation into and out of ArcSDE 8.0 for all of FME's supported formats, which include numerous vector formats. The supported FME formats are constantly evolving. Refer to www.safe.com for a current list. Based on an evaluation of FME (Ruiz and Morrison 2000), FME would be a good tool to reduce the time and manual labor required for MicroStation CAD files to be translated to a GIS format. In addition, FME objects hold promise to provide on-the-fly data conversion in a web application.

Raster Data Storage

Image and raster data are not supported in ArcSDE 8.0, but will be supported ArcSDE 8.1. The prototype system used a beta version of this software and the raster files load properly. Metadata tools for the raster data were not available, however, and that would be troublesome if not resolved before the release. The proposed interim solution is to store physical paths to the raster datasets, thereby facilitating the loading of the data directly from the file, as shown in Figure 11.

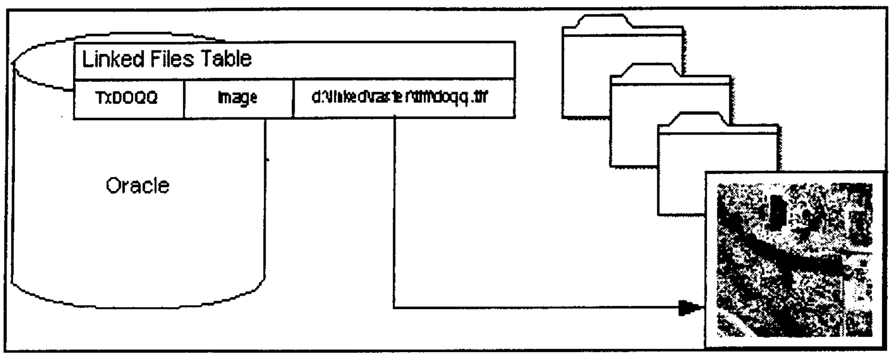


Figure 11. Image storage and retrieval flow.

Versioning

The ArcSDE/Oracle approach to storing geographic data exploits the concepts of object-oriented data models in a relational DBMS environment. Traditional RDBMS data management strengths tend to be focused on the assumption of "short transactions," where multiple and frequent hits on the data base occur in a short period of time. The ability to maintain database stability under such conditions is a hallmark of the modern DBMS. In GIS work flow, however, the tendency is toward "long transactions," where editing may occur over a period of hours or even days. Versioning is a solution in this type of situation, which is supported by ArcSDE (Woodsford 1995; Newell and Batty 1994).

With versioning, a default copy of the data remains intact in the database, while the person with write permission is able to make changes to a different version of the data. Another person with access to that data would not see those updates until the changes were registered. This is different from the short transaction environment, where the row being edited is locked during the short period that the editing occurs. The entire feature class is not duplicated for versioning, the only data being saved for the versioned copy are the new or modified features that have been edited. Multiple users could possibly edit the same feature class

in two different versions. These would then need to be reconciled to the default copy. At the installation, only a limited number of people are able to perform edits on a particular layer, and write permission would be available only with consent of the accountable party. The ability to have more than one version would not likely be necessary.

At the installation, versions of data could be helpful in several areas. Habitat data are often collected when a digging permit is requested in the range area. A biologist may make a visit to field check the available habitat map in areas where field checking had been carried out. Other day-to-day work activities might result in other such updates to the data. It might be preferable to keep one stable version of the habitat map on the DER rather than change it daily, but a different version that reflects the more recent field visits. This newer version could then be used periodically to update the default copy.

Versioning is preferable to making a copy of the data when a feature dataset is being used that includes relationships among feature classes. Using an earlier example, a change to an installation's boundary would result in changes to the training areas adjacent to the boundary. The use of a version in this case makes it possible to keep the relationships intact while the editing is being performed.

4 Web-based Tool for DER

Introduction

The previous chapter discussed various aspects of data storage. This chapter considers three other functions of a data repository. The three functions are:

1. Search – Software issues, user interface, key words, and metadata.
2. Download – Mechanics of data access, form of the data for users and for applications.
3. Upload – Mechanics and personnel needed to get data into the system.

An important component of the proposed DER functionality related to these three areas is a web-based tool that facilitates search, download, and upload functions by setting up a standard manner to handle metadata, by indexing the data in the DER, and by providing a framework for data administration. Recall that the DER may be accessed directly with ESRI tools; the DER web tool offers an alternative. The exception to this is data loading. The web-based tools provide a method to keep the data catalog current and consistent and they provide guidelines for recommended QA/QC (Quality Assurance/Quality Control) procedures as data are added to the DER. This alternative form of access helps keep track of the GIS data in the DER and makes it more accessible to a variety of people.

The major characteristics and benefits of the DER web tool are:

- Upload manager – people who collect new data can submit it to be included in the DER.
- Workflow manager – as data are loaded into SDE, they are cataloged for future reference, checked for proper metadata, and added to the DER data set in a consistent way in accordance with the data accountability responsibilities.
- View maps – data in the DER may be viewed and printed with a web browser.
- Search – data in the DER may be searched according to subject, Spatial Data Standards entity set, or various other parameters.

- Data download – optionally, a function is available so that data can be downloaded in several spatial formats.
- Security management – users are assigned an access type and data layers are assigned accessibility codes in order to protect sensitive data.
- Data integration across the web – map images and data served with web technology are processed differently than accessed directly from the DER. By cross-indexing the data from the DER to the other service, there is a path that keeps track of what data are where and keeps the web-based data current.

Also note that though the data in the DER **may** be made accessible via the Internet, the tool is designed **primarily** as an Intranet tool. This means that the same network capabilities that allow the Internet to work for the world are the same capabilities used inside the installation network.

The web-based tool includes a variety of functions, as listed in Table 11.

Table 11. Web-based tool functions.

Function	Sub-Function	Description
Admin		DER Site Administration Functions
	User Manager	Add Users and Edit User Information; set User permissions to DER site Functions
	LUT Manager	Manage Look Up Tables for selectable metadata elements such as Entity Set, or Catalog Subject
Data Manager		Manage the Data Administrators
	Assignments	View or change the Data Administrator and to whom upload dataset is assigned
Maps		Mapping
	Mapping	Mapping tool to view data and download it
Data Transfer		Upload a dataset to the DER site
	Upload Dataset	Upload a dataset to the DER site
	Modify Upload	Edit the upload information associated with a dataset
Data Admin		Data Administration Functions
	Administer Data	Perform steps to incorporate an uploaded dataset to the DER site
Data Library		Library of information on uploaded datasets
	Search	Search the Data Library for a keyword and download data
Data Manager		Manage the Data Administrators
	Administrators	View the Data Administrators on the DER site; set the Data Administrators Active or Inactive
Help		Help
	Online Help	Access the DER Online Help System
User		User Management
	Change Password	Change your User password

Search

The web-based tool search function allows people to find out what is available in the DER and, optionally to download it (Figure 12). The search is performed on a set of fields stored in an Access database. One of the key features of the search function is the use of the Dublin Core Metadata Initiative (DCMI) standards. The DCMI is a set of 15 elements to describe any type of document or file. It was developed to be a "simple description record for networked resources. It is expected that a simple and widely-understood set of elements will promote interoperability amongst heterogeneous metadata systems and improve resource discovery on the Internet" (Wieble, Innalla, and Cathro 1997).

The 15 elements are described more fully in Appendix C. Each element is described briefly in Table 12. While spatial data in the DER must be accompanied by the FGDC's metadata file, this file is not helpful for search in its current form. FGDC metadata in databases today have usually been created with some metadata tool. The output is a text file and different tools output the data in slightly different forms. The text file format is good for easy transfer of the entire file but is less efficient for creating structure fields of metadata information.

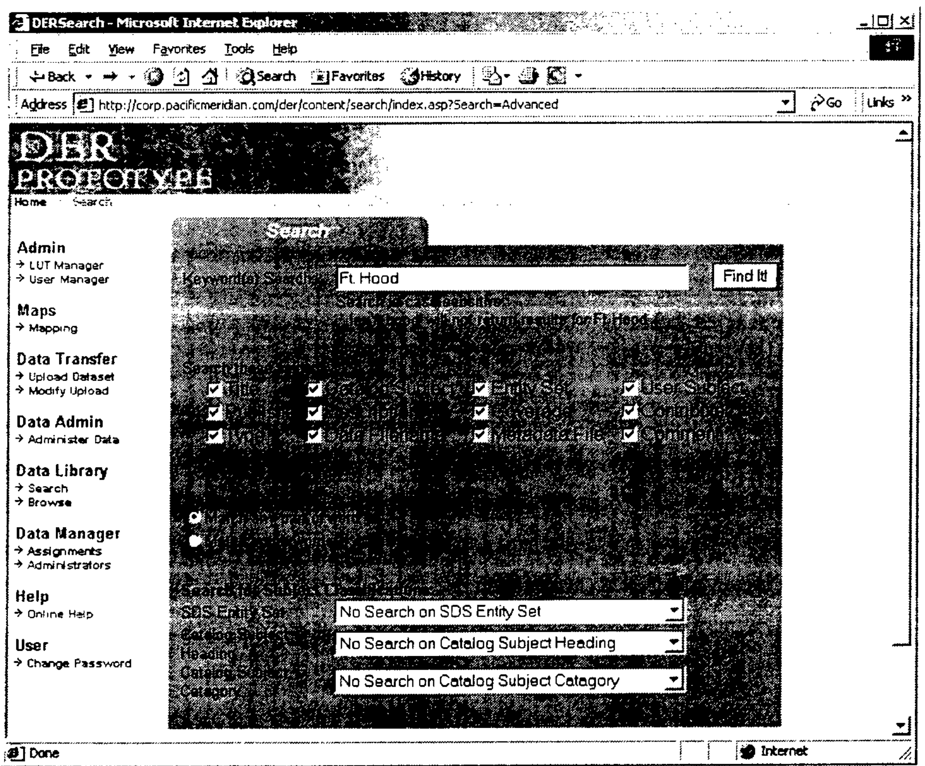


Figure 12. DER search screen.

Table 12. DCMI element description.

DCMI Element	DCMI Meaning	DER Intention
Title	A name given to the resource	FGDC metadata title
Creator	An entity primarily responsible for making the content of the resource	Accountable party as agreed by the stakeholders at Fort Hood
Subject 1	The topic of the content of the resource	Spatial Data Standards Entity Set
Subject 2		Key word – General heading from Fort Hood list
Subject 3		Key word – Secondary heading (Categories) from Fort Hood list
Description	An account of the content of the resource	Free-text account of the content
Publisher	An entity responsible for making the resource available	POC – a named person, not an office
Contributor	An entity responsible for making contributions to the content of the resource	Persons authorized to make changes to the data
Date	A date associated with an event in the life cycle of the resource	Time period of the data
Type	The nature or genre of the content of the resource	Null at this time
Format	The physical or digital manifestation of the resource	Original software format or media type
Identifier	An unambiguous reference to resource within a given context	Unique reference. SDE layer ID or Oracle ID
Source	A reference to a resource from which the present resource is derived	Not in use at this time
Language	A language of the intellectual content of the resource	Null at this time
Relation	A reference to a related resource	Null at this time
Coverage	The extent or scope of the content of the resource	Geographic extent
Rights	Information about rights held in or over the resource	Type of user able to access (search, download, mapping) to information

The Arc8 ArcCatalog metadata tools are available to search on the full FGDC metadata. They seem to be very full-featured, with a lot of opportunity for customization. However, at this time, metadata is stored in a format that is inaccessible except through ArcCatalog. This makes viewing or searching of this data via a web browser impossible. If those metadata files are opened, then a search on the full geographic metadata may be a future option.

In addition to the DCMI elements, the data are indexed on the SDS entity set. This standard is described more fully in Chapter 2. This entity set keyword provides a Corps of Engineers standardized way to reference the theme of data in the DER. Two other keywords are the General Heading and Categories of data established for the assignment of responsibility of the data layers. The category names provided in Chapter 2 are suggestive rather than final and are not exhaustive.

Download and Other Data Access

After searching for data, there is currently an option to download that data from the DER (Figure 13) or to view a map (Figure 14). The data in the DER are all in SDE's format. ArcSDE must thus be engaged to download the data. Currently, the download is accomplished by using Safe Software's FME (discussed in Chapter 3). FME talks to SDE and extracts a file in the format requested. Although FME supports dozens of formats, the current formats available in the DER are shape files, coverages, and DXF (Digital Exchange Format) files.

Two forms of access were described in the preceding paragraph: visualization through mapping and download data files. The access to maps may satisfy most non-GIS users. Core GIS users can connect directly to the DER with a GIS, so they would not necessary require a tool for this. The downloading of data in a GIS or CAD format would be very helpful to people who need to use data for their work, particularly, contractors and Corps of Engineer employees. It may also be useful for GIS or CAD users who are not in the core GIS group.

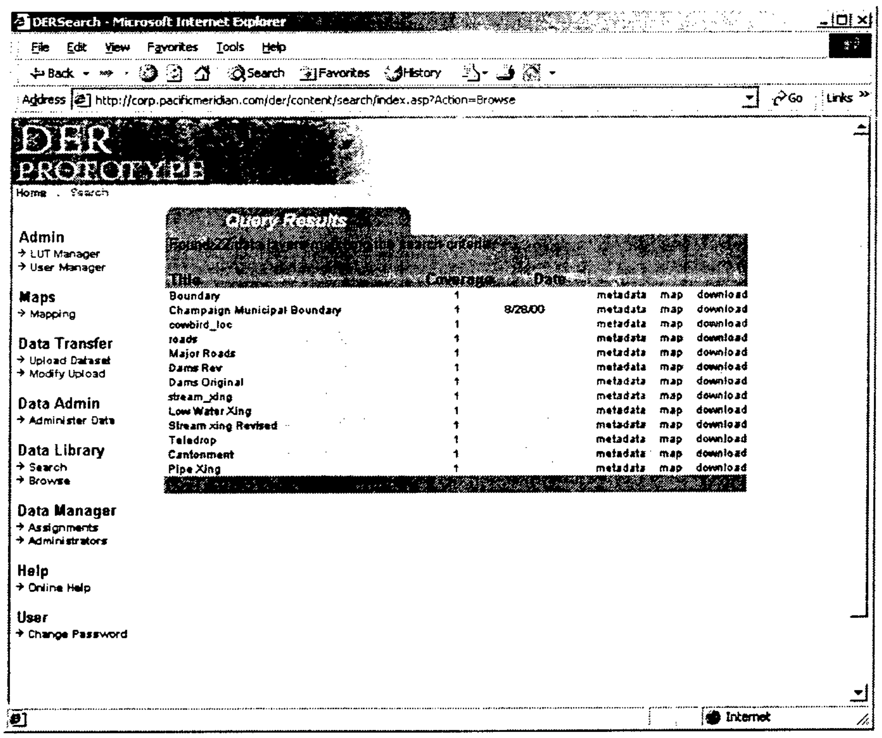


Figure 13. DER download (query results) screen.

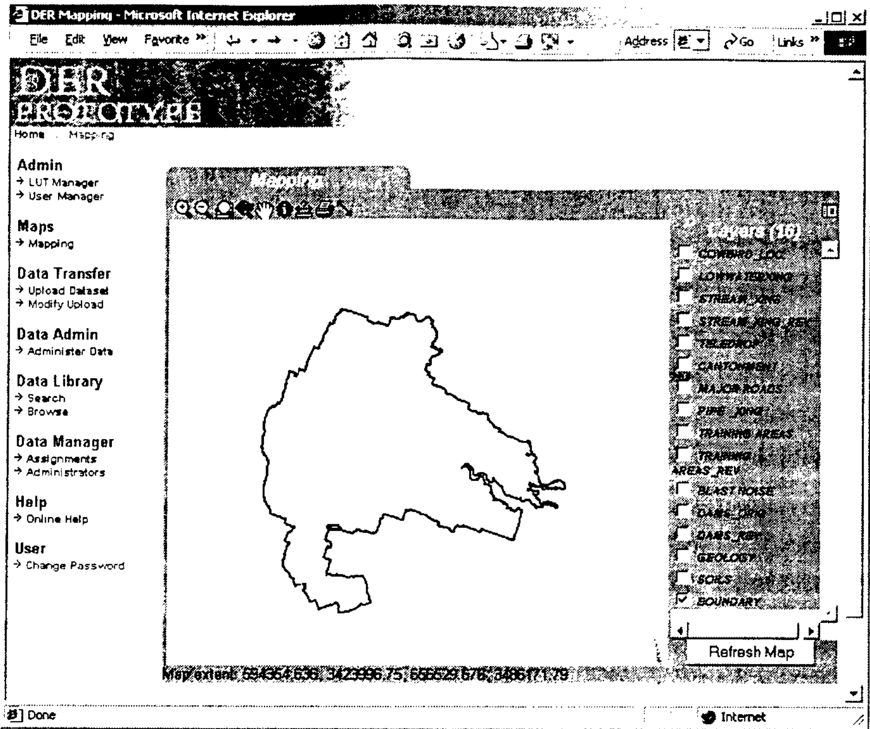


Figure 14. DER view map screen.

Upload

To have data in the DER, somebody has to put it there. The way that this might occur is the topic of this section. The DER web tool includes two important functions for this process. First is a tool that facilitates the delivery of a data file to a data administrator. Second is a work flow manager that guides the administrator through a series of steps to ensure that indexing and quality control occur for every data item in the DER.

Data in the DER may come from a variety of sources. Contractors collect field data on wildlife, for example. Different installation offices create map data for their own purposes, but it may be useful to share that data with others. An academic or government-based research effort might result in data being collected. In these types of cases, the data may be submitted for inclusion to the DER through the data upload tool (Figure 15).

The data upload tool requires a zipped data file and an FGDC-compliant metadata file in text format. In addition, the submitter fills out several fields that are used to identify the content of the data. These fields are somewhat redundant with the FGDC data, but serve as a check to be certain that the FGDC file and the data submitted are consistent with each other. The submit button transfers

the file from the location specified by the submitter to a server specified by the DER web tool program. It also results in an email being sent to a data administrator, who would then be responsible for examining and loading the data into the DER. The data content would dictate to which administrator the notice is sent.

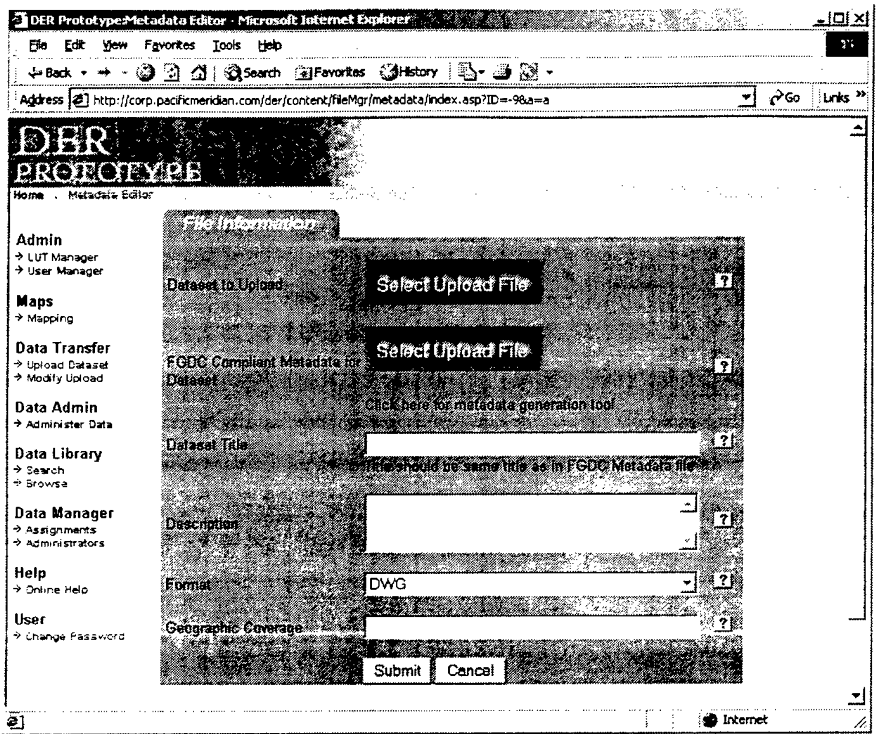


Figure 15. DER data upload screen.

Workflow Manager

The workflow manager guides the data administrator through a series of steps (Figure 16). These steps are defined as:

Step One: Verify composition and completeness of FGDC Metadata. Make sure the metadata is FGDC compliant for the correct dataset and check completeness of the metadata.

Step Two: Verify that the dataset title matches the title in the FGDC metadata. The titles should match exactly and should be unique in the DER.

Step Three: Verify and modify subject fields as necessary. Each of the subject fields should be complete and accurate. Verify input. It is possible to add entries to Entity Sets and Catalog Subjects through a Look Up Tables

(LUT) Manager, also a part of the DER web tool. Make sure any keywords in the narrative field are consistent with other data.

Step Four: Enter the Date of the Dataset (from Metadata). Enter the date from the FGDC metadata. This is used to search datasets on date.

Step Five: Verify Data format. Verify the data format. Add entry to Data Format LUT as necessary using the LUT Manager.

Step Six: Verify and modify source as necessary. Is the data source correct?

Step Seven: Verify and modify coverage as necessary. Choose the dataset geographic coverage from the pull-down list. If the geographic area is not described, add a new Geographic Coverage using the Add button.

Step Eight: Verify and modify access rights. What should the access rights to this dataset be? Verify the setting as necessary.

- Public Access
- General Access
- Secure Access

Step Nine: Incorporate Data Layer into SDE. Follow the installations procedure to upload a dataset into SDE.

Step Ten: Add Layer To MapService if necessary. Add the SDE data layer to the correct ArcIMS MapService using the ArcIMS Manager Tool. Add symbology and scale factors as deemed necessary by the dataset's properties. Add layers to MapServices depending on Rights.

- Public Access: Add to Public MapService
- General Access: Add to General and Public MapServices
- Secure Access: May not be in a MapService

Step Eleven: Input SDE Layer Name. Add the SDE data layer name exactly as it was entered into SDE. Do not use the SDE User name prefix.

This data upload process may seem tedious, but it is a way to slow down the process of data uploading enough to consider what is going into the data set that represents the "official" data for land and cultural resource management at the installation. At the same time, in some cases, where data are being created daily or more frequently, as might occur for monitoring data, it would not be helpful to use this process. A more automated process could be devised that obtained the same set of information without requiring a user/administrator to input it.

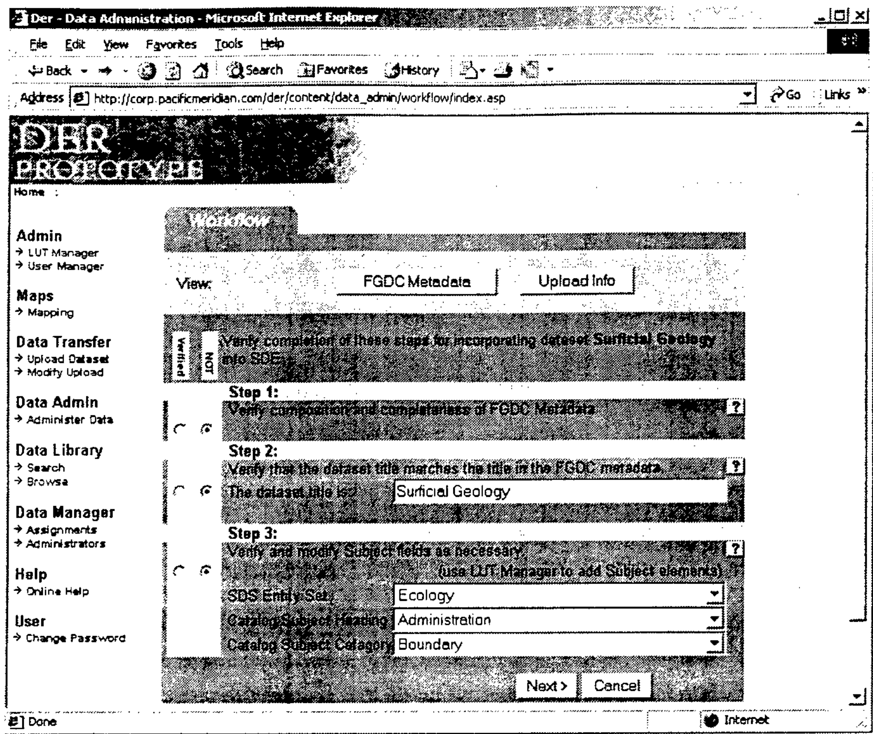


Figure 16. DER workflow manager screen.

Summary

The web-based tool was developed to provide some structure and consistency to the DER data management. While at first sight it appears fairly simple, it offers some fairly complex functionality in a straightforward manner. It is not a fully functional piece of the DER, however, until it is customized and modified to meet site-specific demands. The current tool is distributed with the application so available staff may perform this customization.

A usability evaluation of the DER website showed that it met most typical usability standards. All completed aspects of the site were found to be fully functional and easy to use. National Software Testing Laboratories (NSTL), the usability testers, found the site to be easily navigable with the menus and links being simple to follow and easy to understand. There is consistency in the page design, giving the effect of continuity throughout the site. The site design enables a user to accomplish uploads, downloads, and viewings with very few steps — which enhances its usability. The appearance of pop-up text boxes when the cursor is positioned over a menu item or icon is consistent and useful.

The online help was fairly comprehensive. While the use of screen shots can be useful, NSTL believes adding more descriptive text to complement the use of the

screen shots would provide greater clarity. The help file is currently structured as a single large document. The addition of links, to specific topics, at the top of the document would greatly enhance the speed with which one can access specific information.

NSTL found no facility for feedback to the site administrator on this website. Due to the nature of the site, NSTL recommends the implementation of a link to allow user feedback or comments; email is generally implemented for this purpose.

Overall, the web-based tool provides a possible starting point for an integrated data management plan for geospatial data. It cannot solve all data management problems, but if used consistently, it could help push things in a good direction. The Oracle/SDE approach to storing data will not, by itself, ensure an accurate, up-to-date data store that is readily accessible to all parties who require it. If people are going to discard data that is not the "official" best version of a given layer, they need to be assured that they have a consistent way to get access to the data. This tool helps to carry the potential provided by RDBMS linkage out to providing access to the data.

5 Data Management and Security Issues

In this section, we consider personnel and business aspects of the DER in light of the architecture description provided earlier. This section has two parts: Personnel (administrative roles) and Security (data sensitivity and system encroachment prevention).

Personnel Requirements

The users of the system are described elsewhere. In this section, we consider the support personnel and functions. There are several levels at which this is important.

System Administrator

The system administrator needs to coordinate the activities between Oracle, SDE, and the users of the system. The main responsibilities include:

- Oracle/SDE support – including tuning the SDE database,
- Web application maintenance and enhancement,
- Server administration including passwords, backups, system upgrades, and
- Coordinate with network administrators.

The system administrator role may be shared among a number of people, but it needs to be coordinated by one person to ensure that good performance and ongoing maintenance requirements are addressed. For example, the use statistics of the system will guide the decisions about how to better allocate the table space used by Oracle and SDE.

Data Administrator

The data administrator oversees the loading of data into the Oracle/SDE. This person ensures that QA/QC procedures are being followed and coordinates with the various groups that use the data to be sure that the data and content are up-to-date and new data are being incorporated in a timely manner. The data administrator would also help to coordinate activities related to making more com-

plex and comprehensive data available through the use of the geodatabase capabilities. Primary responsibilities include:

- Process incoming data to the DER,
- Coordinate with the accountable parties to ensure the correct data content and metadata are in the system, and
- Develop plan for more comprehensive data model incorporating behaviors into the data model.

One person should oversee this responsibility, but the duties may be shared among the core GIS users at the installation. For example, the duties may be fulfilled through:

- ITAM GIS analyst,
- GIS support for Natural Resources, or
- GIS support for Environmental Division/Cultural Resources.

Network Administration

The Directorate of Information Management (DOIM) maintains the network at the installation. There may be some need for further internal data security measures, as outlined in the next section. The System Administrator would need to work with DOIM to ensure that network support is optimal and all security measures are enacted.

Security and Data Sensitivity

When data are available on a networked system, there is concern about the ability to ensure that the data are not misused or that sensitive data might fall into the hands of people who would use it in destructive ways. Several security issues have arisen in discussions about the DER. The main concerns are that:

1. Sensitive data would be made available
   a. Locations of archaeological sites would be made available to potential looters and treasure hunters
   b. Cave locations would be made available and endanger the sensitive habitat there
2. The physical vulnerability of the network might be exploited by hackers through the public access portion of the DER. This would allow a breach in the security of the data on the network at large.
3. The centralized storage paradigm makes it more possible for more people to get access to data that could be misinterpreted. Map data are commonly several de-

grees of abstraction away from the original field source. The proper interpretation of the data may not be possible without appropriate professional judgment.

4. Litigation or sanctions against the installation based on misinterpreted information could arise more readily due to the relatively easy access of the data by the public.

These issues have played an important role in guiding the design of the web application for data access and the recommended procedures for determining DER content. The following several strategies were used in helping to alleviate concerns and protect data from misuse.

1. The DER is largely an Intranet, rather than an Internet resource. The same technologies for transferring data are at work for both Intranet and Internet systems, but the Intranet portion of the system is protected by the network infrastructure and firewalls in place at the installation. The intention of the system is to provide a centralized data store to a defined set of people – not to the world.
2. For the portion of the system that serves data to the "world," a well-defined set of data is all that is available. The server that provides that information via the Internet is separate from the server that holds the core DER content. That portion, and only that portion, of the system operates outside the firewall.
3. The web-based application is designed to allow the system administrator to set the type of access to data and to application functionality allowed for any given user. In addition, the data administrators assign each data layer a value in terms of what type of user is allowed to have access to it. The types of users are:
    a. System Administrator – Full access to entire system
    b. Data Administrator – Access to most functionality and all data
    c. General User On-Site – Access to base functionality and most data
    d. General User Off-Site – Access to very restricted functionality and a set of data
    e. Research/Special User – Specific functionality and specific data that as required by the task
4. During the data inventory, certain data were identified as being sensitive. Each one needs to be handled in accordance with the judgment of the accountable party for the layer (see Table 7).

Security Assessment

To provide a comprehensive look at existing vulnerabilities, an independent security assessment was carried out by the Science Applications International Corporation's (SAIC) Secure Business Solutions Group (SBSG). The general re-

sults of the report are discussed here. The specific results are not included, but were made available to installation personnel for internal use.

SAIC's Center for Information Security Technology's Information Security Assessment (ISA) approach identified threats to Pacific Meridian's developmental DER system and data; discovered vulnerabilities by way of external attacks of the system; and assessed the adequacy of the security infrastructure in place at the proposed location of the production DER.

The SAIC study proceeded through the following stages as SAIC:

- studied available DER documentation to extract security issues and concerns.
- examined existing network resources to identify the underlying security functionality required to protect them.
- conducted external network penetration testing in order to identify existing vulnerabilities in the development system.
- examined the security features of the products being used to provide the DER functionality in order to determine recommended settings and identify security weaknesses.
- conducted a walk-through inspection of the proposed production site to identify physical security issues and concerns.
- interviewed network and administration personnel to determine if appropriate security procedures have been established.
- worked with key DER stakeholders in defining threats and in valuing critical information assets.
- determined network risks and possible countermeasures.
- produced a summary ISA report that documents the findings and recommendations specific to the DER system.

A systematic threat model was used to evaluate the data system (Figure 17). Interviews with key stakeholders and technical personnel indicate the greatest perceived threat lay with outsiders gaining access to cultural information maintained in the DER system, as well as hackers, including those persons possessing authorized access to the installation LAN (Local Area Network) who seek to compromise other systems on the LAN. As the DER has not yet been placed into operation as a production system, the following threats are anticipated based on our experience.

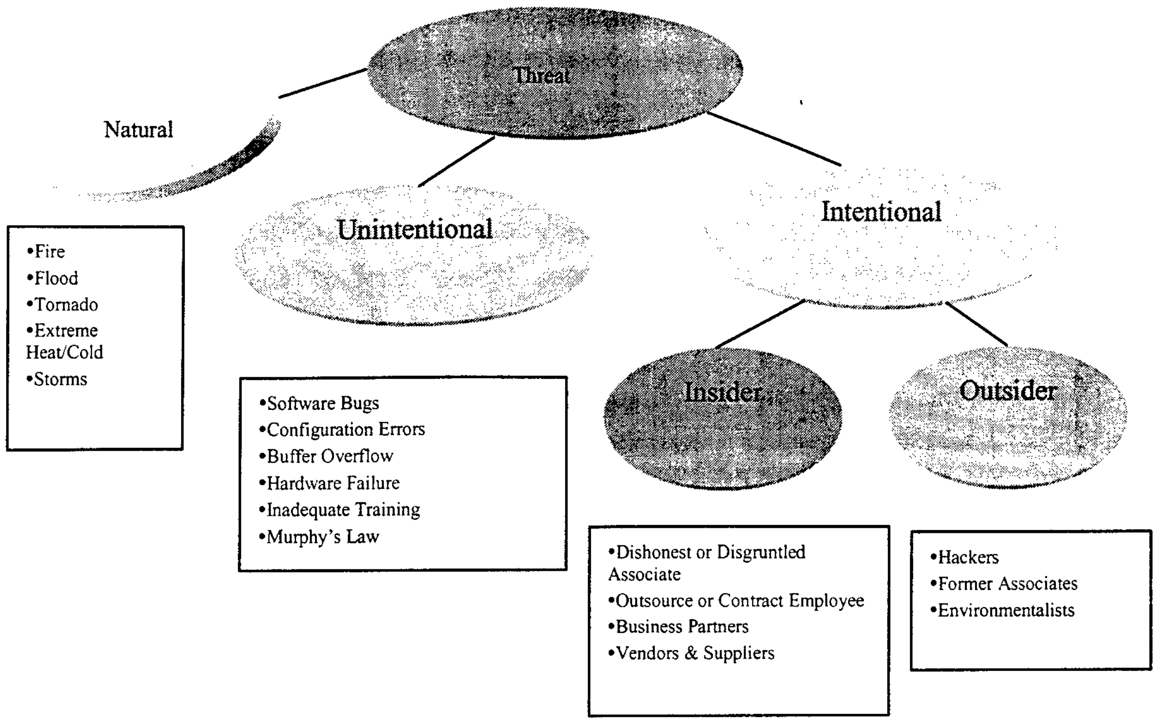


Figure 17. Systematic threat model.

Intentional Threats from Man

We focused the bulk of our efforts on the harm one or more individuals might cause if they deliberately set out to harm the DER through misuse or damage to information systems. Intentional threats can be divided into two broad categories: inside people and outside people.

Inside People. Permanent or temporary users, system administration personnel, contractors, and authorized LAN users with malicious intent pose the biggest threat to security simply by virtue of their knowledge of the systems. Typically, they have the most knowledge about the operation and the most access to the premises. They have an association with the government/Army that can be a source of irritation to them. Their behavior can result in the widest range of possible scenarios with the potential to harm the DER.

Visitors allowed physical access to equipment, especially if not observed, could steal, sabotage or damage out of ignorance. Frequent visitors can gain significant knowledge about operations in seemingly casual conversation with local personnel or contractors. Vendors and suppliers who enter the site should be treated as frequent visitors. Personnel should be reminded that although they are familiar faces or voices on the telephone, they are not entitled to sensitive information.

Outside People. Former users, system administrators, and maintenance contractors pose the second biggest threat to security. They have detailed knowledge of operations from their employment and may have grudges against former co-workers, supervisors, or managers. Hackers who already have access to the installation LAN also pose significant threat. Due to the lax internal controls of the installation LAN, and the frequency of security incidences experienced on the network, hacking is a major concern. Also of concern is the outside hacker (non-authorized installation LAN user). The degree to which hackers pose a threat to the DER will be largely based on the technical controls and system configurations in place when the DER is put into production.

Persons motivated by anger or rage, which can include former employees, may engage in retaliatory actions such as denial of service attacks and/or attempts to destroy DER resources. Persons motivated by political differences may attempt to obtain/modify sensitive DER information in an effort to help their cause. Persons motivated by greed can engage in espionage, sabotage, fraud, and theft of information. Espionage can result in loss or disclosure of sensitive information. Persons motivated by curiosity or challenge may destroy or deny access to DER data and systems just for the thrill of it.

Unintentional Threats from Man

A wide variety of harms can result from human negligence or accident. Installing new data equipment exposes operational equipment to construction/installation accidents, misconfiguration, or unanticipated delays. Most likely harms are power and communications interruptions and equipment "infant mortality" failures during or shortly following installation.

External to the sites, a construction accident could cut cables causing power or communications loss. We did not inquire whether communications circuits were diversely routed (reducing the chance of an errant backhoe cutting all communications into and out of building 4216).

Unintentional misuses of data systems, out of ignorance or by accident, present some risk of disruption to DER operations. IT system complexity makes it nearly impossible to foresee the consequences of every possible mistake humans can make. Careful system design, back-up and error checking processes, and good "user friendly" human-computer interface standards are useful in reducing mistakes and reducing the potential for damage from any single mistake.

Threats from Nature

We assessed the highest likelihood natural threat in the installation area to be storm/tornado. Tornadoes have the potential to seriously disrupt DER operations. Beyond direct damage to facilities, tornadoes have the potential to disrupt the area's power grids and hamper personnel attempting to reach work.

Flooding may be possible, but we did not see any evidence that raised major concern. Fire is always a possibility and major concern. Fire may be caused by natural events or man made circumstances.

Recommendations and Findings

Information Security Policy

The number one recommendation from the SAIC report is the need for well-defined and broadly distributed security policy, both for the installation and for the DER. An Information Security Policy is the foundation for all aspects of security within the network. The policy identifies at the highest level "what needs protection," and "whom should it be protected from." The security policy should state the philosophy of protection, objectives, rules, and practices with respect to security and the control of how resources (including information) are managed, protected, and distributed within the network and at external interfaces. It should be well integrated with physical security and personnel management processes.

Information Security Policy objectives, among others, are put into effect through guidelines and procedures. These guidelines and procedures are typically developed to address specific aspects of the network or business process and can consist of both technical and nontechnical measures. A high-level security policy currently does not exist for the installation LAN. A policy is said to be under development. Additionally, while preliminary DER security issues have been considered (i.e., sensitivity of data, possible role-based access profiles), security policy, guidelines, and procedures for operation of the production DER do not exist.

All the formal and informal guidelines and procedures form the defacto business process. That process must have a cornerstone in overall risk management. Otherwise, the individual processes of well-meaning managers can leave security gaps. Users and network administrators achieve (or fail to achieve) security goals by how well (or poorly) guidelines and procedures support security objec-

tives. Publishing these guidelines and procedures establishes a basis for internal dialogs, security audits, and reasonable risk management.

Fort Hood Network Architecture Analysis

It was found that the security from outside the installation is well developed. The barriers in place provide a defense in depth, which, in the opinion of SAIC would be very difficult for even an expert hacker to surmount, assuming that these defenses are properly configured and maintained. The more important vulnerability came within the installation. More attention needs to be given to the potential of damage from users within Fort Hood. Some straightforward remediation would greatly improve the situation.

DER Web-based Application

Numerous vulnerabilities were found at the Pacific Meridian site. Most were related to the administration of the site, itself. Care should be taken upon implementation that those potential vulnerabilities are not transferred to the installation.

Overall, the potential threats to the system were related to management issues. Remediation of these issues is highly site specific and is outside the scope of this report. At the same time, they should not be dismissed from a study of technology. The implementation of a system such as the proposed DER is on the border between technology and business processes and the two pieces cannot be taken in isolation if a system that works is to be available at the installation.

6 Summary and Conclusions

The data management conditions at the installation are stable. An incremental approach to DER implementation would provide the least disruptive path to improving the data sharing and access. The direction for the next set of activities needs to be organized by the primary users. Items to be considered should include the following.

1. Determine responsibility for making certain that the accountability requirements are agreeable and executable for all involved. Individuals need to set up plans to execute this responsibility as needed.
2. Determine logistics for providing the correct data layers for input to the DER. These data must be in proper form and be accompanied by FGDC documentation.
3. Make a plan for the fuller adoption of the SDS for the standard data layers in the DER.
4. The options for a strong organizational element to provide on-going coordination of GIS activities.

Additional recommended actions include:

- FME from Safe Software, Inc, may facilitate the translation of data from the Bentley MGE format to an ESRI coverage format. This would involve setting up a custom "mapping file" that carries over attributes and spatial entities from the CAD to the GIS file formats. This is not the black box approach that is available with many GIS import/export tools. With FME, you can create a custom configuration that not only extracts specific features from the source data, but can also assign additional attributes to the data or change the spatial type of the feature extracted to a different type.
- When the accountable parties determine the content and supply the data files for loading into the DER, they should also ensure that old versions and copies of the data are no longer present on the servers and workstations. This will help to address the data integrity issues and will wipe the slate clean for future data management.
- The data in the DER will not be tiled and it will be in one spatial reference system. The spatial reference system will be WGS84 and UTM meters. ArcSDE stores data in its own coordinate space. The implications of how this is

set up need to be discussed and agreed upon before loading data into the DER. The GIS core group needs to ensure that the coordinate space in the DER is suitable for their needs.

- A part of the implementation process needs to include a plan for a security policy, along with express guidelines and procedures for operation of the full-production version of the DER. This must be tailored to reflect the decisions about how to implement the DER.
- A follow-up security assessment should be undertaken following implementation of the DER.

For the DER architecture, the hardware and software in the summary in Table 10 include all software for both the core DER and the web-based access. Without the web-based access tool, fewer components are necessary. Some form of the functionality in the web-based tool is highly recommended, but a much simpler, less cohesive system could be developed without it.

For geometry storage, the SDE binary format is the most straightforward. Future requirements where data need to be accessed from some other application may point to the need for one of the other geometry options.

This report does not make recommendations for the precise SDE coordinate information and spatial index characteristics. These need to be made when the final decisions are made about which data are to be included in the DER.

The object-oriented geodatabase holds great promise for a more structured and well designed data set. Its complexity, however, could overshadow other immediate implementation tasks. For this reason, it should be considered part of a future phase in the DER.

The web-based tool is considered a prototype. It delivers considerable functionality, but it needs to be tailored to installation personnel and requirements to be as useful as possible.

The security assessment pointed to good control with regard to Internet access to data. This should allay the fears of those who are concerned about hackers getting into data from outside the installation.

The policies and practices inside the installation's computer network are considerably more vulnerable to foul play. The existing data on the various GIS computers are currently vulnerable. A change to a more centralized system will reduce the possibility of data misuse. It is important to remove the various

datasets found on the current workstations after they are loaded into the DER so that the geospatial data are less vulnerable to network weaknesses.

At the time of this writing, the DER is not in place at the testing site installation. Before a specific plan can be developed to move forward toward a new data management and storage system, the stakeholders must evaluate the proposed plan, taking into account the various aspects of the proposed system. The details in this report come from a fairly complex set of issues from the testing site installation, from technology trends, and from the LMS research directive. The parts that were particular to the installation were an effort to build on the existing system, while taking into account the difficulties with that system and the expressed needs of the individual users. The outcome of that evaluation is in the hands of the installation stakeholders.

References

ArcSDE Configuration and Tuning Guide for Oracle. 1999. Redlands, CA: ESRI Press.

Booch, Grady, James Rumbaugh, and Ivar Jacobson. 1999. *The Unified Modeling Language User Guide*. Reading, MA: Addison-Wesley.

CAD/GIS Technology Center, http://tsc.wes.army.mil/products/TSSDS-TSFMS/tssds/html/.

Campus Glossary. 2000. ESRI, Inc.'s Campus Web Page (http://campus.esri.com/glossary/geodbase.cfm).

Directorate of Public Works. 2000. *U.S. Army, Fort Hood, Texas, Web Page* (http://www.dpw.hood.army.mil). Rhonda Michael, Web Master.

Fowler, Martin and Kendall Scott. 1999. *UML Distilled*. 2nd ed. Reading, MA: Addison-Wesley.

Freeman, Edward. 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

GIS Hydro 2000 Pre-Conference Seminar. 2000. *ArcGIS Hydro Data Model*. Unpublished Manuscript (http://utwired.engr.utexas.edu/crwr/cd_Consortium_2000/introduction/intro2000.html).

Goran, William D., Jeffery P. Holland, John W. Barko, and Andrew J. Bruzewicz. 1999. *Plans for the Land Management System (LMS) Initiative*. ERDC/CERL Technical Report 99/60/ADA366731 (June). Champaign, IL.

Gorman, Michael M. March 1999. "Great News – The Relational Model is Dead!" The Data Administration Newsletter @ http://www.tdan.com/i008hy02.htm. www.tdan.com.

Harris, Mark and Jillian Clark. 1999. *ArcSDE Administration Guide*. Redlands, CA: ESRI Press.

MacDonald, Andrew. 1999. *Building a Geodatabase*. Redlands, CA: ESRI Press.

Newell, Richard G. and Peter M. Batty. 1994. "GIS Databases are Different." *AM/FM International*, 279-288.

Ruiz, Marilyn and Dawn A. Morrison. 2000. "FME 2000: Translation With Oomph!." First Impressions Column, *Geospatial Solutions* (October 2000).

Understanding ArcSDE: Integrating GIS and RDBMS. 1999. Redlands, CA: ESRI Press.

Wieble, Stuart, Renato Innalla, and Warwick Cathro. 1997. *User Guidelines for Dublin Core Creation*. Version 1.0. Paper generated as part of the 4th Dublin Core Metadata Workshop Re-

port, National Library of Australia, Canberra (continuously reviewed and improved, see http://www.sics.se/~preben/DC/DC_guide.html).

Woodsford, Peter. 1995. "The Significance of Object-Orientation for GIS." Laser Scan Limited Technology Papers (http://www.laserscan.com/papers/). Based on a paper first presented at the IUSM Conference, Hannover, Germany, September 1995.

Zeiler, Michael. 1999. Modeling Our World: The ESRI Guide to Geodatabase Design. Redwoods, CA: ESRI Press.

Acronyms

API	Application Programming Interface
ASP	Active Server Pages
BLOB	Binary Large Object
CAD	Computer-Aided Design
COM	Component Object Model
CRMT	Cultural Resources Management Team
DBA	Database Administrator
DBMS	Database Management System
DCMI	Dublin Core Metadata Initiative
DD	Decimal Degrees
DEM	Digital Elevation Model
DENTAC	Dental Activity
DER	Data Enterprise Repository
DHTML	Dynamic HyperText Markup Language
DOIM	Directorate of Information Management
DOQQ	Digital Orthographic Quarter Quadrangle
DPW	Directorate of Public Works
DLG	Digital Line Graph
DRG	Digital Raster Graphics
DXF	Digital Exchange Format
EMB	Environmental Management Branch
EMT	Energy Management Team
ENG	Engineering Division
ENV	Environmental Division
EPS	Engineering Plans and Services
ERDC	Engineer Research and Development Center
ESRI	Environmental Systems Research Institute
FGDC	Federal Geographic Data Committee
FME	Feature Manipulation Engine
G3	Range Safety Office/Range Engineering Office
GIS	Geographic Information System
GRASS	Geographic Resources Analysis Support System
HTML	HyperText Markup Language
IIS	Internet Information Server
IMS	Internet Mapping System

ISA	Information Security Assessment
ITAM	Integrated Training Area Management
LAN	Local Area Network
LMS	Land Management System
LUT	Look Up Table
MEDDAC	Medical Support Activity
MTS	Microsoft Transaction Server
NIMA	National Imagery and Mapping
NRB	Natural Resource Branch
NSTL	National Software Testing Laboratories
OO	Object Oriented
POC	Point of Contact
QA/QC	Quality Assurance/Quality Control
RB	Recycling Branch
RDBMS	Relational Database Management System
SAIC	Science Applications International Corporation
SBSG	Secure Business Solutions Group
SDE	Spatial Database Engine
SDEAPI	Spatial Database Engine Application Programming Interface
SDS	Spatial Data Standard
SDS FMS	Spatial Data Standards and Facility Management Standards
SQL	Standard Query Language
TEXCOM	Test and Experimentation Command
TIN	Triangulated Irregular Network
TNC	The Nature Conservancy
USACE	U.S. Army Corps of Engineers
UTM	Universal Transverse Mercator

Appendix A: SDS Entity Set Definitions

Entity Set Name	SDS Code	Definition
Boundary	bd	The borders or boundaries that define logical or political divisions or subdivisions
Buildings	bg	The structures located on the face of the earth that were created, by man, to protect man and his possessions from the environment; or to enhance man's activities
Cadastre	cd	The man-made division of land into areas of ownership and control
Common	cm	The information that describes the overall data set or components of data that are common to all entity sets
Communications	co	The means available to relay data and translate data
Cultural	cr	The activities of man that are historically significant
Demographics	de	The information pertaining to man's trends or traits
Ecology	ec	The information pertaining to the interrelationship between organisms and their environments
Environmental Hazards	eh	The identification and management of natural and manmade substances, materials, and conditions which are, or have the potential to be, detrimental to life and ecosystems on the earth
Fauna	fa	The study of the animals in a region or environment
Flora	fl	The study of the plant life in a region or environment
Future Projects	fp	The information that describes planned projects for future development
Geodetic	gd	The information pertaining to the size and shape of the earth
Geology	ge	The geologic features and processes occurring in a given region on the earth
Hydrography	hy	The physical conditions, boundaries, flow, and related characteristics of earth's waters
Improvement	im	The miscellaneous man-made minor structures and facilities which improve appearance, provide security, or facilitate man's activities
Land Status	ls	The current use by man of the surface of the earth
Landform	lf	The distribution of features that make up the visible surface of the earth's crust
Military Operations	ml	The information relevant to military presence, operations, training, and security
Transportation	tr	The methods and means of spatial movement in a large scale
Utilities	ut	The man-made components of a system that provides a service to the public. The components of each utility system in this entity set are located outside of the foundation of a structure. Communication systems are not included and constitute a separate entity set

Appendix B: Potential Duplicate/ Redundant Data Themes

For more information on each record, please see full data spreadsheet. Dark lines delineate each group of potentially duplicate/redundant data.

Sheet #	Simple Name	Location Code
59	Unsurveyed archeological sites	DPW-2
68	Unsurveyed Archaeology Areas	DPW-2
115	area designations?	DPW-2
131	areas of jurisdiction	DPW-3
90	Army Corps land	DPW-2
133	Army Corps land	DPW-3
116	buildings	DPW-2
143	buildings	DPW-3
213	burn data (vegetation)	TNC2
208	burned to not burned hab?	TNC2
113	caves	DPW-2
129	caves	DPW-3
11	digsites	ITAM-1
88	digsites	DPW-2
42	DOQQ	ITAM-1
252	DOQQs	TNC2
180	GCWA Banding Locations, 1991-1998	TNC1
182	GCWA Banding Locations, 1991-1996	TNC1
1	Fort Hood & Surrounding Area Maps	ITAM-1
146	area maps (counties, TX, Ft. Hood boundary)	DPW-3
19	Geology & Soils	ITAM-1
156	geology and soils (Ft Hood & TX)	DPW-4
16	Infrastructure of Installation	ITAM-1
155	features (airfield, building, roads, sidewalks, structures)	DPW-4
160	land descriptions; cross country military maneuvers	DPW-4
36	Landform & Topography	ITAM-1
6	Live fire	ITAM-1
132	outline of live fire area	DPW-3
209	Live fire area Mask	TNC2
64	live fire training areas	DPW-2

30	Railroads	ITAM-1
121	railroads	DPW-2
166	railroads	DPW-4
178	road data	DPW-4
69	Roads	DPW-2
33	Roads-miscellaneous	ITAM-1
37	Texas	ITAM-1
249	Texas data	TNC2
163	various data pertaining to the state of Texas	DPW-4
3	Training Areas	ITAM-1
62	Training areas	DPW-2
39	Utilities	ITAM-1
170	Utilities (detailed)	DPW-4
164	waterbodies	DPW-4
159	waterbodies; various ponds & lakes	DPW-4
21	Hydrology	ITAM-1
67	watersheds	DPW-2
196	Watersheds (small)	TNC2
157	watersheds; small & large	DPW-4

Appendix C: DCMI Elements Descriptions

Description of document:

This document is the reference description, version 1.1 of the Dublin Core Metadata Element Set. This document supersedes the Dublin Core Metadata Element Set, version 1.0. See the Dublin Core Home Page (http://dublincore.org/) for further information about the workshops, reports, working group papers, projects, and new developments concerning the Dublin Core Metadata Element set.

Document Metadata: http://dublincore.org/documents/1999/07/02/dces

Introduction:

The document summarizes the updated definitions for the Dublin Core metadata elements as originally defined in [RFC2413]. These new definitions will be officially known as Version 1.1.

The definitions utilize a formal standard for the description of metadata elements. This formalization helps to improve consistency with other metadata communities and enhances the clarity, scope, and internal consistency of the Dublin Core metadata element definitions.

Each Dublin Core element is defined using a set of ten attributes from the ISO/IEC 11179 [ISO11179] standard for the description of data elements. These include:

- **Name** - The label assigned to the data element
- **Identifier** - The unique identifier assigned to the data element
- **Version** - The version of the data element
- **Registration Authority** - The entity authorized to register the data element
- **Language** - The language in which the data element is specified

- **Definition** - A statement that clearly represents the concept and essential nature of the data element
- **Obligation** - Indicates if the data element is required to always or sometimes be present (contain a value)
- **Datatype** - Indicates the type of data that can be represented in the value of the data element
- **Maximum Occurrence** - Indicates any limit to the repeatability of the data element
- **Comment** - A remark concerning the application of the data element

Fortunately, six of the above ten attributes are common to all the Dublin Core elements. These are, with their respective values:

Version: 1.1

Registration Authority: Dublin Core Metadata Initiative

Language: en

Obligation: Optional

Datatype: Character String

Maximum Occurrence: Unlimited

The above attributes will not be repeated in the below definitions, however, they do represent part of the formal element definitions.

The definitions provided here include both the conceptual and representational form of the Dublin Core elements. The Definition attribute captures the semantic concept and the Datatype and Comment attributes capture the data representation.

Each Dublin Core definition refers to the resource being described. A resource is defined in [RFC2396] as "anything that has identity." For the purposes of Dublin Core metadata, a resource will typically be an information or service resource, but may be applied more broadly.

Element: Title

Name:	Title
Identifier:	Title
Definition:	A name given to the resource.
Comment:	Typically, a Title will be a name by which the resource is formally known.

Element: Creator

Name:	Creator
Identifier:	Creator
Definition:	An entity primarily responsible for making the content of the resource.
Comment:	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.

Element: Subject

Name:	Subject and Keywords
Identifier:	Subject
Definition:	The topic of the content of the resource.
Comment:	Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Element: Description

Name:	Description
Identifier:	Description
Definition:	An account of the content of the resource.
Comment:	Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

Element: Publisher

Name:	Publisher
Identifier:	Publisher
Definition:	An entity responsible for making the resource available
Comment:	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.

Element: Contributor

Name:	Contributor
Identifier:	Contributor
Definition:	An entity responsible for making contributions to the content of the resource.
Comment:	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.

Element: Date

Name: Date
Identifier: Date
Definition: A date associated with an event in the life cycle of the resource.
Comment: Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format.

Element: Type

Name: Resource Type
Identifier: Type
Definition: The nature or genre of the content of the resource.
Comment: Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the working draft list of Dublin Core Types [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.

Element: Format

Name: Format
Identifier: Format
Definition: The physical or digital manifestation of the resource.
Comment: Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware, or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).

Element: Identifier

Name: Resource Identifier
Identifier: Identifier
Definition: An unambiguous reference to the resource within a given context.
Comment: Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

Element: Source

Name:	Source
Identifier:	Source
Definition:	A Reference to a resource from which the present resource is derived.
Comment:	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

Element: Language

Name:	Language
Identifier:	Language
Definition:	A language of the intellectual content of the resource.
Comment:	Recommended best practice for the values of the Language element is defined by RFC 1766 [RFC1766] which includes a two-letter Language Code (taken from the ISO 639 standard [ISO639]), followed optionally, by a two-letter Country Code (taken from the ISO 3166 standard [ISO3166]). For example, 'en' for English, 'fr' for French, or 'en-uk' for English used in the United Kingdom.

Element: Relation

Name:	Relation
Identifier:	Relation
Definition:	A reference to a related resource.
Comment:	Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

Element: Coverage

Name:	Coverage
Identifier:	Coverage
Definition:	The extent or scope of the content of the resource.
Comment:	Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range), or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and that, where appropriate, named places or time periods be used in preference to numeric identifiers such as sets of coordinates or date ranges.

Element: Rights

Name:	Rights Management
Identifier:	Rights
Definition:	Information about rights held in and over the resource.
Comment:	Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.

References for Appendix C

[DCT1] List of Resource Types. Dublin Core Draft Working Group Report. <http://dublincore.org/DC/documents/1999/08/05/resource-typelist

[ISO11179] ISO 11179 - Specification and Standardization of Data Elements, Parts 1-6. <ftp://sdct-sunsrv1.ncsl.nist.gov/x318/11179/>

[ISO639] ISO 639 - Codes for the representation of names of languages. <http://www.oasis-open.org/cover/iso639a.html>

[ISO3166] ISO 3166 - Codes for the representation of names of countries. <http://www.oasis-open.org/cover/country3166.html>

[MIME] Internet Media Types. <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types

[RFC1766] Tags for the Identification of Languages, Internet RFC 1766. <http://www.ietf.org/rfc/rfc1766.txt>

[RFC2396] Uniform Resource Identifiers (URI): Generic Syntax, Internet RFC 2396. <http://www.ietf.org/rfc/rfc2396.txt>

[RFC2413] Dublin Core Metadata for Resource Discovery. Internet RFC 2413. <http://www.ietf.org/rfc/rfc2413.txt>

[TGN] Getty Thesaurus of Geographic Names. <http://www.getty.edu/research/tools/vocabulary/tgn/index.html

[W3CDTF] Date and Time Formats, W3C Note. <http://www.w3.org/TR/NOTE-datetime>

CERL Distribution

Chief of Engineers
ATTN: CEHEC-IM-LH (2)

Commander, Fort Hood
ATTN: Environmental Division (3)

Engineer Research and Development Center (Libraries)
ATTN: ERDC, Vicksburg, MS
ATTN: Cold Regions Research, Hanover, NH
ATTN: Topographic Engineering Center, Alexandria, VA

Defense Tech Info Center 22304
ATTN: DTIC-O

9
6/00

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)
05-2001

2. REPORT TYPE
Final

3. DATES COVERED (From - To)

4. TITLE AND SUBTITLE
Geospatial Data Enterprise Repository: A Report on the Prototype for Fort Hood, Texas

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Marilyn Ruiz, Dawn Morrison, David Bouwman, Kevin McNinch, Frank Schreiner, and Shari Forbes

5d. PROJECT NUMBER
622720A917

5e. TASK NUMBER

5f. WORK UNIT NUMBER
BH0

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
P.O. Box 9005
Champaign, IL 61826-9005

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/CERL TR-01-46

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander U.S. Army Corps of Engineers
441 G Street, NW
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)
CERD-ZA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Copies are available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

14. ABSTRACT
Geographic Information System (GIS) data management options have expanded rapidly over the past several years. Software and hardware advances have provided better network access to spatial data, allowed more complex geospatial data models, and have made the integration of GIS and database management systems a reality.

At Fort Hood, Texas, GIS technology has been used extensively for military land management. Though their current technology is mature, there remain several issues related to geospatial data management that hinder efficiency, including: duplication of data themes, uncertainty about accountability for the content of key data themes, turn over of GIS staff with a subsequent loss of institutional knowledge, and time-consuming data requests.

This report outlines a proposed plan for a centralized geospatial data repository, the Data Enterprise Repository (DER), to meet geospatial data requirements. The core users of the system are the Environmental Division Office, Cultural Resource Management Team, Natural Resource Branch, and the Integrated Training Area Management office. The DER will also benefit those who need to view maps from the system but do not use a GIS, and will enable people offsite, who are performing work at the installation, to get better access to the data they require for their work.

15. SUBJECT TERMS

Ft. Hood, TX	geospatial data	data management
Data Enterprise Repository (DER)	geographic information system (GIS)	

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Marilyn Ruiz
Unclassified	Unclassified	Unclassified	SAR	84	19b. TELEPHONE NUMBER (include area code) (217) 352-6511, ext 7368